Comput Optim Appl (2009) 43: 235–259
DOI 10.1007/s10589-007-9136-7

Mean and variance optimization of non–linear systems and worst–case analysis

P. Parpas · B. Rustem · V. Wieland · S. Žaković

Received: 8 June 2006 / Revised: 9 July 2007 / Published online: 9 November 2007

Abstract In this paper, we consider expected value, variance and worst–case optimization of nonlinear models. We present algorithms for computing optimal expected value, and variance policies, 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 expected performance in a stochastic setting.

Both approaches are used with a small macroeconomic model to illustrate relative performance, robustness and trade-offs between the alternative policies.

Keywords Expected value · Worst–case analysis · Policy design

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 shock, 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

P. Parpas · B. Rustem (✉) · Žaković
Department of Computing, Imperial College, London SW7 2AZ, UK
e-mail: br@doc.ic.ac.uk

V. Wieland
Goethe University Frankfurt, Frankfurt am Main, Germany

expected value evaluation used in [15, 18] for systems governed by parametrized feedback rules. The starting point for the worst–case optimization approach is that of Rustem and Howe [17]. 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.

An alternative approach to worst–case robust design is the H^∞ strategy [1]. The latter approach augments the objective function with a concave term. We have followed the more intuitive approach based on a continuum of scenarios contained in a compact set.

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.

The contribution of this paper is the development of an algorithmic framework for the expected value optimization of nonlinear systems. Moreover, we compare the expected value approach to that of worst–case analysis. The computational comparison is made on a real world application from economics. 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. The computational experiments illustrate the efficacy of the two frameworks as well as their limitations.

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 $F(x, v)$:

$$\min_x \mathrm{Var}_v(F(x, v)). \tag{4}$$

The motivation for minimizing the variance of F stems from the application oriented requirement of computing strategies that do not deviate far from the expected performance. In (3) we minimize the deviation of the L_2 norm of the functions from zero, whereas in (4) we minimize the deviation from the mean performance of the system. Similar concepts are used in the design of minimum variance controllers in control theory.

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 [3, 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, 15]. The following two results are standard, we will use them repeatedly in the sequel, and so we state them as propositions.

Proposition 2.1 *Let $\epsilon \in R^n, \epsilon \sim \mathcal{N}(0, \Lambda)$, and $Q \in R^{n\times n}$ a symmetric matrix. Then we have*

$$E(\epsilon^T Q\epsilon) = \mathrm{trace}(\Lambda Q).$$

Proposition 2.2 *Let $v \in R^n, v \sim \mathcal{N}(\overline{v}, \Lambda)$, and $q \in R^n$. Then*

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

Proposition 2.3 *Let $v \in R^n, v \sim \mathcal{N}(0, \Lambda)$ and Q a symmetric matrix of dimension n. Then*

$$E_v[(v^T Qv)]^2 = [\mathrm{trace}(\Lambda Q)]^2 + 2\,\mathrm{trace}(\Lambda Q)^2.$$

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:

$$E(bb^T) = I, \qquad (v^T Qv) = (b^T Bb),$$

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 [16] that

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

Consider the transformed expression:

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

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

- $i = j,\ k = l,\ i \neq k$;
- $i = k,\ j = l,\ i \neq j$;
- $i = l,\ j = k, \dot{\imath} = j$;
- $i = j = k = l$.

So we have

$$\begin{aligned} E(b^T Bb)^2 &= \sum_{i,k,i\neq k} B_{ii} B_{kk} + \sum_{i,j,i\neq j} B_{ij}^2 + \sum_{i,j} B_{ij} B_{ji} + 3\sum_i B_{ii}^2 \\ &= \sum_{i,k} B_{ii} B_{kk} + 2\sum_{i,j} B_{ij}^2 = [\text{trace}(B)]^2 + 2\,\text{trace}(B^2). \end{aligned} \tag{5}$$

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

$$\begin{aligned} &\text{trace}(B) = \text{trace}(\Lambda^{1/2} Q \Lambda^{1/2}) = \text{trace}(\Lambda^{1/2}\Lambda^{1/2} Q) = \text{trace}(\Lambda Q), \\ &\text{trace}(B^2) = \text{trace}(\Lambda^{1/2} Q \Lambda^{1/2} \Lambda^{1/2} Q \Lambda^{1/2}) = \text{trace}(\Lambda Q \Lambda Q) = \text{trace}(\Lambda Q)^2. \quad \square \end{aligned}$$

2.1 Expected value optimization

A naive approach to solve (3) is to use a standard nonlinear programming algorithm, and perform function evaluations and gradient estimations using a numerical integration routine. However, performing numerical integration is time consuming, and such an approach will not be applicable to large-scale problems. In this paper we propose to solve problem (3) by using a Taylor series expansion in the neighborhood of $\overline{v}$. The motivation for using a Taylor series expansion is that the integration can be carried out analytically. No doubt this approximation introduces some error into the problem, we then proceed to find an estimate of this error and take it into account in the next iteration.

A second order expansion, with respect to the random variables, is used for approximating the original problem:

$$\widehat{f_i}(x, v) \triangleq f_i(x, \overline{v}) + \nabla_v f_i(x, \overline{v})^T (v - \overline{v}) + \frac{1}{2}(v - \overline{v})^T \nabla_v^2 f_i(x, \overline{v}) f_i(v - \overline{v}),$$

where $\nabla_v f_i(x, \overline{v})$, and $\nabla_v^2 f_i(x, \overline{v})$ denote the gradient and Hessian of f_i respectively. They are both evaluated at $(x, \overline{v})$. We then evaluate the error term, $\hat{\alpha}_i(x)$, such that the equality

$$E_v(f_i(x, v)^2) = E_v(\widehat{f_i}(x, v) + \hat{\alpha}_i(x))^2, \tag{6}$$

is satisfied. Therefore $\hat{\alpha}_i(x)$ are (possibly complex conjugate) roots of the quadratic equation:

$$E_v(\widehat{f_i}(x,v)^2) + 2\hat{\alpha}_i(x)E_v(\widehat{f_i}(x,v)) + \hat{\alpha}_i(x)^2 - E_v(f_i(x,v)^2) = 0, \tag{7}$$

with real valued coefficients.

At the k^{th} iteration the algorithm proceeds to calculate the next point x_{k+1} as follows: for fixed x in (6) to x_k, estimate $\hat{\alpha}_i(x_k)$ using Monte Carlo simulation. Then using the latter error estimate, the right hand side of (6) is minimized to obtain x_{k+1}. For the minimization problem to be efficiently solved we need to be able to compute, in closed form, the expectation in the r.h.s. of (6). This calculation can be easily performed as follows:

$$\begin{aligned}
&E_v(\widehat{f_i}(x,v)^2) \\
&= E_v(f_i(x,\overline{v}) + \nabla_v f_i(x,\overline{v})^T(v-\overline{v}) + \frac{1}{2}(v-\overline{v})^T\nabla_v^2 f_i(x,\overline{v})(v-\overline{v}))^2 \\
&= E_v(f_i(x,\overline{v})^2) + E_v(\nabla_v f_i(x,\overline{v})^T(v-\overline{v}))^2 \\
&\quad + 2f_i(x,\overline{v})E_v(\nabla_v f_i(x,\overline{v})^T(v-\overline{v})) \\
&\quad + \frac{1}{4}E_v((v-\overline{v})^T\nabla_v^2 f_i(x,\overline{v})(v-\overline{v}))^2 \\
&\quad + E_v((f_i(x,\overline{v}) + \nabla_v f_i(x,\overline{v})^T(v-\overline{v}))((v-\overline{v})^T\nabla_v^2 f_i(x,\overline{v})(v-\overline{v}))) \\
&= f_i(x,\overline{v})^2 + \nabla_v f_i(x,\overline{v})^T\Lambda\nabla_v f_i(x,\overline{v}) + \frac{1}{4}(\text{trace}(\Lambda\nabla_v^2 f_i(x,\overline{v})))^2 \\
&\quad + \frac{1}{2}\text{trace}(\Lambda\nabla_v^2 f_i(x,\overline{v}))^2 + f_i(x,\overline{v})\text{trace}(\Lambda\nabla_v^2 f_i(x,\overline{v})).
\end{aligned}$$

At the lth iteration, the algorithm proceeds by solving the following optimization problem:

$$\min_x \sum_{i=1}^{k} E_v(\widehat{f_i}(x,v)^2) + 2\hat{\alpha}_i(x_{l-1})E_v(\widehat{f_i}(x,v)) + \hat{\alpha}_i(x_{l-1})^2, \tag{8}$$

in order to obtain x_l. In (8) only the first two terms are actually used, the last term is constant in the lth iteration and can be ignored. The calculation of the expectations of the first two terms can easily be performed as described above.

We now turn our attention to the calculation of the $\hat{\alpha}_i$ in (7). Solving (7) is straightforward once $E_v(f_i(x,v)^2)$ is known. We estimate the latter quantity using a Monte Carlo method. Let $\{v_j\}_{j=1}^{N(l)}$ denote $N(l)$ i.i.d samples from a Gaussian density with mean $\overline{v}$ and variance Λ. In (7) we use the following estimate:

$$E_v(f_i(x,v)^2) \approx \frac{1}{N(l)}\sum_{j=1}^{N(l)} f_i(x,v_j)^2.$$

The above estimate is arbitrarily good provided that $N(l)$ is large enough. Thus at the lth iteration, x_{l-1} is available, and the $\hat{\alpha}_i$'s are chosen to satisfy:

$$E_v(\widehat{f_i}(x_{l-1}, v)^2) + 2\hat{\alpha}_i(x_{l-1})E_v(\widehat{f_i}(x_{l-1}, v)) + \hat{\alpha}_i(x_{l-1})^2 - \frac{1}{N(l)}\sum_{j=1}^{N(l)} f_i(x_{l-1}, v_j)^2 = 0. \tag{9}$$

We note that if $f_i(x, v)$ is quadratic in v, then $\hat{\alpha}_i(x) = 0$, 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 the above 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 $\hat{\alpha}_i(x)$, 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: $l = 1$, choose x_0.
STEP 1: Calculate $\hat{\alpha}_i(x_l)\ \forall i$, using MC simulation (see (9)).
STEP 2: Solve (see (8)):

$$x_{l+1} \in \arg\min_x E_v(F(x, v)).$$

STEP 3: Check for convergence: if $\frac{\|x_{l+1}-x_l\|}{\|x_l\|} \leq \epsilon$ stop, otherwise $l = l + 1$, goto STEP 1.

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 [15, 18], even for nonlinear models.

In order to account for the error introduced by the Monte Carlo simulation we will need the following result. It is a simple generalization of the weak law of large numbers.

Proposition 2.4 *Let* $\{x_l\}$ *be any sequence converging to* x, *and suppose that* $\lim_{l\to\infty} N(l) = \infty$. *Furthermore suppose that*:

$$\mathrm{Var}(f_i(x, v)) < \infty \quad \forall x,\ i = 1, \ldots, m.$$

Then given any $\epsilon > 0$,

$$\lim_{l\to\infty} P\left(\left|\frac{1}{N(l+1)}\sum_{j=1}^{N(l+1)} f_i(x_l, v_j)^2 - E_v(f_i(x, v)^2)\right| > \epsilon\right) = 0.$$

Proof Using our assumption that $\mathrm{Var}(f_i(x,v)) \le c_1 < \infty$, it follows by the dominated convergence theorem (see e.g. [4]) that:

$$\lim_{l\to\infty} E_v(f_i(x_l,v)^2) = E_v \lim_{l\to\infty} (f_i(x_l,v)^2) = E_v(f_i(x,v)^2).$$

It follows that there exists a $L > 0$, such that

$$|E_v(f_i(x_l,v)^2) - E_v(f_i(x,v)^2)| \le \frac{\epsilon}{2} \quad \forall l \ge L.$$

Therefore,

$$\begin{aligned}
&\left| \frac{1}{N(l+1)} \sum_{j=1}^{N(l+1)} f_i(x_l,v_j)^2 - E_v(f_i(x,v)^2) \right| \\
&= \left| \frac{1}{N(l+1)} \sum_{j=1}^{N(l+1)} f_i(x_l,v_j)^2 - E_v(f_i(x_l,v)^2) \right. \\
&\quad \left. + E_v(f_i(x_l,v)^2) - E_v(f_i(x,v)^2) \right| \\
&\le \left| \frac{1}{N(l+1)} \sum_{j=1}^{N(l+1)} f_i(x_l,v_j)^2 - E_v(f_i(x_l,v)^2) \right| \\
&\quad + \left| E_v(f_i(x_l,v)^2) - E_v(f_i(x,v)^2) \right| \\
&\le \left| \frac{1}{N(l+1)} \sum_{j=1}^{N(l+1)} f_i(x_l,v_j)^2 - E_v(f_i(x_l,v)^2) \right| + \frac{\epsilon}{2}.
\end{aligned}$$

Using the preceding estimate we obtain:

$$\begin{aligned}
&P\left(\left| \frac{1}{N(l+1)} \sum_{j=1}^{N(l+1)} f_i(x_l,v_j)^2 - E_v(f_i(x,v)^2) \right| > \epsilon \right) \\
&\le P\left(\left| \frac{1}{N(l+1)} \sum_{j=1}^{N(l+1)} f_i(x_l,v_j)^2 - E_v(f_i(x_l,v)^2) \right| + \frac{\epsilon}{2} > \epsilon \right) \\
&= P\left(\left| \frac{1}{N(l+1)} \sum_{j=1}^{N(l+1)} f_i(x_l,v_j)^2 - E_v(f_i(x_l,v)^2) \right| > \frac{\epsilon}{2} \right). \qquad (10)
\end{aligned}$$

Using the Chebychev [4] inequality we obtain:

$$P\left(\left|\frac{1}{N(l+1)}\sum_{j=1}^{N(l+1)} f_i(x_l,v_j)^2 - E_v(f_i(x_l,v)^2)\right| > \frac{\epsilon}{2}\right)$$
$$\leq \frac{4}{N(l+1)^2\epsilon^2}\mathrm{Var}\left(\sum_{j=1}^{N(l+1)} f_i(x_l,v_j)^2\right) \leq \frac{4c_1}{N(l+1)\epsilon^2}.$$

Consequently:

$$\lim_{l\to\infty} P\left(\left|\frac{1}{N(l+1)}\sum_{j=1}^{N(l+1)} f_i(x_l,v_j)^2 - E_v(f_i(x_l,v)^2)\right| > \frac{\epsilon}{2}\right) = 0.$$

The result now follows from (10). □

Proposition 2.5 *In addition to the assumptions of Proposition 2.4, suppose that $\{x_l\}$ is a sequence generated in Step 2 of Algorithm 1, and that it remains in a compact set $\mathcal{X}$. Then since $\mathcal{X}$ is compact the sequence $\{x_l\}$ generated by the algorithm has a limit point, and any such point minimizes the expectation of $F(x,v)$ on $\mathcal{X}$.*

Proof Let x be an arbitrary point in $\mathcal{X}$. Let $\{x_l\}$ be any sequence converging to x. At every step of the algorithm the following function is minimized:

$$\sum_{i=1}^{k} c_i(x,l) \triangleq \sum_{i=1}^{k} E_v(\widehat{f_i}(x,v) + \hat{\alpha}_i(l))^2,$$

where $\hat{\alpha}_i(l)$ satisfy:

$$E_v(\widehat{f_i}(x_{l-1},v)^2) + 2\hat{\alpha}_i(l)E_v(\widehat{f_i}(x_{l-1},v)) + \hat{\alpha}_i(l)^2$$
$$-\frac{1}{N(l)}\sum_{j=1}^{N(l)} f_i(x_{l-1},v_j)^2 = 0, \quad i = 1,\ldots,k.$$

Evaluating the limit in the expression above, we have that:

$$\lim_{l\to\infty} E_v(\widehat{f_i}(x_{l-1},v)^2) + 2\hat{\alpha}_i(l)E_v(\widehat{f_i}(x_{l-1},v)) + \hat{\alpha}_i(l)^2$$
$$= \lim_{l\to\infty}\frac{1}{N(l)}\sum_{j=1}^{N(l)} f_i(x_{l-1},v_j)^2.$$

Using Proposition 2.4:

$$\lim_{l\to\infty}\frac{1}{N(l)}\sum_{j=1}^{N(l)} f_i(x_{l-1},v_j)^2 = E_v(f_i(x,v)^2) \quad \text{i.p.}$$

Consequently we must have that:

$$\lim_{l\to\infty} E_v(\widehat{f_i}(x_{l-1},v)^2) + 2\hat{\alpha}_i(l)E_v(\widehat{f_i}(x_{l-1},v)) + \hat{\alpha}_i(l)^2 = E_v(f_i(x,v)^2) \quad \text{i.p.}$$

Summing over all the k functions we obtain:

$$\lim_{l\to\infty}\sum_{i=1}^{k} c_i(x_l,l) = \lim_{l\to\infty}\sum_{i=1}^{k} E_v(\widehat{f_i}(x_l,v) + \hat{\alpha}_i(l))^2 = \sum_{i=1}^{k} E_v(f_i(x,v)^2).$$

The preceding equation establishes that the sequence $\{\sum_{i=1}^{k} c_i(x_l,l)\}_l$, converges in probability to the original function. Convergence in probability is stronger than convergence in distribution. Therefore we have also shown the sequence $\{\sum_{i=1}^{k} c_i(x_l,l)\}_l$, converges in distribution to the original function. It follows from Theorem 2.8 in [2] that the sequence also epi-converges to the original function. It follows from Theorem 2(b) in [12] that if x^* is an accumulation point of the sequence $\{x_l\}$ generated by Algorithm 1, then x^* is a solution to the original problem. Such an accumulation point exists because $\mathcal{X}$ is compact. Furthermore, if $\{x_{l(v)}\}$ is any subsequence of $\{x_l\}$ converging to x^* we have:

$$\lim_{v\to\infty} \min \sum_i c_i(x_{l(v)}, l(v)) = \min \sum_i E_v(f_i(x,v)^2). \qquad \square$$

Convergence can only be established along a subsequence of $\{x_l\}$ since the objective of the problem is assumed to have multiple minima.

The conditions in Proposition 2.5 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:

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

where the mean $\overline{v}$ may be zero.

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

$$g(x) = \mathbf{1}_{\{x\in K\}}(x-\mu), \tag{11}$$

where:

$$\begin{aligned}
\mathbf{1}_{\{x\in K\}} &= \begin{cases} 1 & \text{if } x \in K, \\ 0 & \text{otherwise,} \end{cases} \\
\mu_i &= \int_K \omega_i n(\omega) d\omega, \qquad (12) \\
n(\omega) &= \frac{\exp(-\frac{1}{2}\omega^T \Lambda^{-1}\omega)}{(2\pi)^{\frac{n}{2}} det(\Lambda)^{\frac{1}{2}}}.
\end{aligned}$$

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:

$$\mathrm{Var}_v(F(x,v)) = E_v[F(x,v) - E_v(F(x,v))]^2. \tag{13}$$

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

$$f(x,v) = \begin{bmatrix} f_1(x,v) \\ f_2(x,v) \\ \vdots \\ f_k(x,v) \end{bmatrix}. \tag{14}$$

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

$$\begin{aligned} &\widehat{f}_i^l(x,v) = f_i(x,\overline{v}) + \nabla_v f_i(x,\overline{v})^T(v-\overline{v}), \\ &E(f_i(x,v)) = f_i(x,\overline{v}). \end{aligned} \tag{15}$$

As in previous section, $\delta_i(x)$ will represent the expected deviation of $\mathrm{Var}(f_i(x,v)^2)$ and $\mathrm{Var}(\widehat{f}_i^l(x,v)^2)$. $\delta_i(x)$ is calculated so that it satisfies:

$$\mathrm{Var}(f_i(x,v)^2) = \mathrm{Var}(\widehat{f}_i^l(x,v) + \delta_i(x))^2, \tag{16}$$

where $\delta_i(x)$ is estimated using a quasi–Monte Carlo simulation. The variance term in the r.h.s. of (16) is evaluated as follows:

$$\begin{aligned} \mathrm{Var}_v(\widehat{f}_i^l(x,v)^2) &= E\left(\nabla_v f_i(\overline{v})^T\epsilon\right)^4 + 4f_i(x,v)^2\nabla_v f_i(\overline{v})^T\Lambda\nabla_v f_i(\overline{v}) \\ &\quad - \left(\nabla_v f_i(\overline{v})^T\Lambda\nabla_v f_i(\overline{v})\right)^2 \\ &= [\mathrm{trace}(\Lambda Df_i(x))]^2 + 2\,\mathrm{trace}(\Lambda Df_i(x))^2 \\ &\quad + 4f(x)^2\nabla_v f_i(\overline{v})^T\Lambda\nabla_v f_i(\overline{v})^T - (\nabla_v f_i(\overline{v})^T\Lambda\nabla_v f_i(\overline{v}))^2, \end{aligned} \tag{17}$$

where $Df_i(x) = \nabla_v f_i(\overline{v})^T\nabla_v f_i(\overline{v})^T$. The problem of minimizing the variance is formulated as:

$$\min_x \mathrm{Var}_v(F(x,v)) = \min_x \sum_{i=1}^{k} \mathrm{Var}_v(\widehat{f}_i^l(x,v) + \delta_i(x))^2. \tag{18}$$

If $f_i(x, v)$ is linear in v then $\delta_i(x) = 0$, and $\mathrm{Var}_v(F(x, v))$ is exactly computed by:

$$\mathrm{Var}_v(F(x, v)) = \sum_{i=1}^{k} \mathrm{Var}_v(\widehat{f_i}(x, v)^2).$$

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

3 The minimax approach

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.

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

$$\begin{aligned} G_1 = \min_x \quad & \max_v \quad F(x, v), \\ & \text{s.t.} \quad \overline{v} - \Delta \le v \le \overline{v} + \Delta, \quad \Delta > 0. \end{aligned} \tag{19}$$

Due to the hypercube constraining (19), the problem above is referred to as box-constrained. Robustness and the price paid for this desirable property, has been the topic of interest for a number of years [11]. Robustness is ensured by an optimality condition. Let x^*, v^* solve (19). Then we have

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

Let

$$\Phi(x) = \max_{\overline{v}-\Delta \le v \le \overline{v}+\Delta} F(x, v), \tag{20}$$

for all x. We call $\Phi(x)$ the max–function. Therefore, (19) can be written as

$$\min_x \Phi(x). \tag{21}$$

To solve (21) 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 (20) as well as a step size strategy that ensures sufficient decrease in $\Phi(x)$ at each iteration (the algorithm is described in [20]).

Problem (21) 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 (21) 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 [17, 19, 22].

There is an interesting relationship between the order of the min and max operators in (19) that is worth discussing. This relationship can best be explained when one attempts to view (19) as a two–person zero–sum game between the Decision Maker (DM) and Nature (N).

In this game player (DM) chooses the x variables, and player (N) chooses the v variables. If (DM) chooses x', and (N) chooses v', then (DM) pays $F(x', v')$ to (N). Naturally, player (DM) wishes to minimize this quantity, while player (N) attempts to maximize it.

In game theory equilibria play an important role. An equilibrium, in the present context, means a point from which no player will gain by a unilateral change of strategy. For the game outlined above an equilibrium point (x^*, v^*) must satisfy:

$$F(x, v^*) \geq F(x^*, v^*) \geq F(x^*, v) \quad \forall x \in X, \ \forall v \in [\overline{v} - \Delta, \overline{v} + \Delta]. \tag{22}$$

A point satisfying the preceding equation is also known as a saddle point of F. Consider the following situation: Player (DM) chooses a strategy first, and then player (N) chooses a strategy. Thus (N) already knows the strategy that (DM) has chosen. As a result (N) will have an advantage. Player (DM) will argue as follows: "If I choose x, then (N) will choose $\max_v F(x, v)$, therefore I better choose the strategy that will minimize my losses." In other words (DM) will choose the optimal strategy given by solving (19).

Now consider the same game, but with the order of play reversed, i.e. Player (N) chooses first, and Player (DM) second. Then applying the rules of rational behavior (as above) we see that (N) will select the v that solves:

$$G_2 = \max_{v \in [\overline{v} - \Delta, \overline{v} + \Delta]} \min_x F(x, v). \tag{23}$$

In our modeling framework it is thus natural to consider the game in (19) where the (DM) decides first and (N) second, since uncertainty is in the future. It is well known that the game will have a saddle point if F is convex in x and concave in v. Under such convexity assumptions then $G_1 = G_2$. Unless convexity is assumed it is difficult to know whether the order of the min and max operators can be changed.

A different way of formulating (19) would be to allow the distribution of v to be uncertain. For example, v can be assumed to be a random variable with mean $\overline{v}$ and variance belonging to the set $[\overline{v} - \Delta, \overline{v} + \Delta]$. The minimax problem can then be

formulated as follows:

$$\begin{aligned} \min_{x} \quad & \max_{\mu \in \mathcal{P}(K)} \quad \int F(x, v) d\mu(v) \\ & \text{s.t.} \quad \int v d\mu(v) = \mu, \\ & \qquad \overline{v} - \Delta \leq \int (v - \overline{v})^2 d\mu(v) \leq \overline{v} + \Delta. \end{aligned}$$

Where $\mathcal{P}$ denotes the set of probability measures with their supported contained in some compact set K. The formulation above is just one possible way to view the minimax problem in a different context. For numerical methods dealing with such problems we refer the interested reader to [5, 13].

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:

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

- Minimization of expected value performance:

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

where Δ is a positive scalar parameter. In the minimax model, Δ has the effect of enlarging the space from which worst case scenarios can be constructed. In the expected value model it represents volatility. The algorithm proposed in this paper will be used in the next section to analyze an economic model. Before we delve into that model, we show in Fig. 1 how the bias of the algorithm converges. It can be seen from Fig. 1 that the bigger the volatility the bigger the distance between to successive iterates. However, the algorithm seems to be robust for different values of Δ. The exact details of the model solved to obtain the results shown in Fig. 1 will be described in the next section.

4.1 A model of the economy

In a recent paper, Orphanides and Wieland [14] use a simple macroeconomic model of inflation, output and interest rates to investigate different motives for inflation point

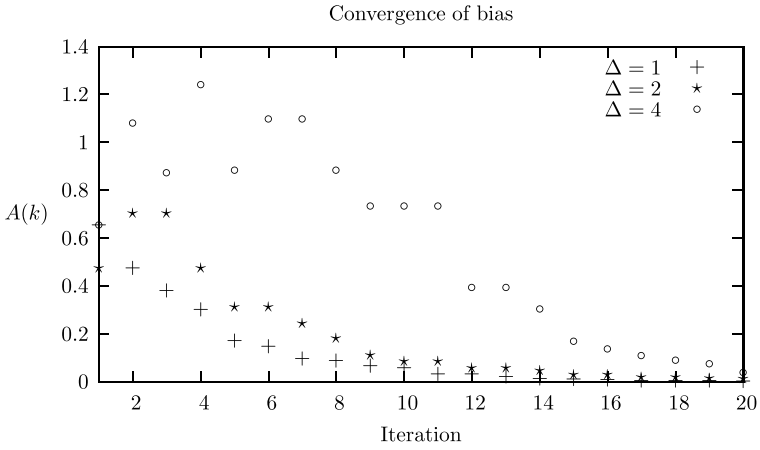

Fig. 1 Convergence of algorithm. $A(k)^2 = \sum_i (\hat{\alpha}_i(x_k) - \hat{\alpha}_i(x_{k-1}))^2$

versus inflation zone targeting. In the first case, the policymaker 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 [14]). 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 [14]).

In the minimalist macro model of [14], 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 the rest of this Section we will first set up the model as a dynamic optimization model, we will then describe its solution using a feedback rule.

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:

$$\min_r E\left\{\sum_{t=1}^{\infty} \beta^{t-1} l_t\right\}. \tag{24}$$

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.

$$l_{t+1} = w(\pi_{t+1} - \pi^*)^2 + (1-w)y_{t+1}^2, \quad w \in (0,1), \tag{25}$$

where w is a weighting parameter. The following two equations describe the evolution of the economy:

$$\begin{aligned} y_t &= (\rho + \overline{\rho})y_{t-1} - \xi r_t + \delta + u_t, \\ \pi_t &= \pi_{t-1} + (\alpha + \overline{\alpha})(\rho + \overline{\rho})y_{t-1} - (\alpha + \overline{\alpha})\xi r_t + (\alpha + \overline{\alpha})\delta + (\alpha + \overline{\alpha})u_t + e_t, \end{aligned} \tag{26}$$

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

$$\alpha, \rho, u_t, e_t \sim \mathcal{N}(0, \Lambda), \quad \forall t. \tag{27}$$

As shown in [14], the optimal policy incorporates a target zone for inflation if the Phillips curve is nonlinear. This is the case, if small output deviations have a more than proportional impact on inflation. This can be modeled by means of a zone–generating function $Z(\cdot\,;\zeta\,;c)$ of the following form:

$$Z(x;\zeta;c) = x - \frac{1}{2}\sqrt{c + \left(x + \frac{\zeta}{2}\right)^2} + \frac{1}{2}\sqrt{c + \left(-x + \frac{\zeta}{2}\right)^2}. \tag{28}$$

The motivation behind the definition of Z is given in [14]. In the numerical examples in this paper we use $c = 0.1$. ζ is a parameter controlling the width of the zone generated by the function. Using the zone–generating function Z, the following equations can be used to describe the evolution of the economy:

$$\begin{aligned} y_t &= (\rho - \overline{\rho})y_{t-1} - \xi r_t + \delta + u_t, \\ \pi_t &= \pi_{t-1} + (\alpha + \overline{\alpha})Z(y_t;\zeta;c) + e_t. \end{aligned} \tag{29}$$

When constraints (26) are used in the model, the model is referred to as the *point targeting model*. When (29) are used, the resulting model is referred to as the *zone targeting model*. We note that in both cases the proposed model is non–linear (and non quadratic) in the uncertain parameters, and therefore the generality of the algorithm described in Sect. 2.1 is required in order to solve the expected value model.

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

$$F(r, v) = \sum_{t=1}^{\infty} \beta^{t-1} l_t. \tag{30}$$

In order to solve the model we use an approach described in Tetlow and von zur Muehlen [21]. In this approach (also Hansen and Sargent in [9], see [10] for a related

approach) the policymaker chooses the parameters x_1 and x_2 of the feedback law:

$$r_t = x_1 \pi_{t-1} + x_2 y_{t-1}. \tag{31}$$

This rule is referred to as a feedback rule, and the x's are referred to as the feedback parameters. The problem can be formulated as:

$$\min_{x_1, x_2} E_v(F(x, v)), \tag{32}$$

where E_v denotes the expectation computed for the uncertain variables v, the objective function F is given by (30). The constraints on the systems are given by (26) or (29). The feedback law is given by (31). Thus the decision we are seeking are the values of the two parameters x_1 and x_2, and these two parameters are valid for all time periods. In this framework, the decision maker commits to following the feedback rule in the future. Let

$$f_1(x, v) = \begin{bmatrix} \beta^{\frac{1}{2}}(\pi_1 - \pi^*) \\ \beta^{\frac{2}{2}}(\pi_2 - \pi^*) \\ \vdots \\ \beta^{\frac{T}{2}}(\pi_T - \pi^*) \end{bmatrix}, \qquad f_2(x, v) = \begin{bmatrix} \beta^{\frac{1}{2}} y_1 \\ \beta^{\frac{2}{2}} y_2 \\ \vdots \\ \beta^{\frac{T}{2}} y_T \end{bmatrix}, \tag{33}$$

then, the objective function can be formulated as:

$$F(x, v) = w f_1^T f_1 + (1 - w) f_2^T f_2, \tag{34}$$

so the problem becomes:

$$\min_x \{w E_v(f_1^T(x, v) f(x, v)) + (1 - w) E_v(f_2^T(x, v) f_2(x, v))\}. \tag{35}$$

4.2 Computational experiments

The results that follow are obtained for $\pi* = 2$, $T = 20$, $\beta = 0.9$ 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 [14] 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 bounds on the uncertainties. For example, by $\frac{1}{2}\sigma_\alpha$ we mean that the α variable was allowed to vary in the set $[\mu_\alpha - \frac{1}{2}\sigma_\alpha, \mu_\alpha - \frac{1}{2}\sigma_\alpha]$. The two objective functions, F_p and F_z correspond to the point and zone models respectively. By F_t^p we denote:

$$F_t^p = w\beta^t(\pi_t - \pi^*)^2 + (1 - w)\beta^t y_t,$$

so that $F^p = \sum_{t=1}^{T} F_t$, similar notation is used for the zone model. As was to be expected the objective function increases as the bounds on the uncertainty increase. In fact as Fig. 2 illustrates, the objective function value increases exponentially with the

Table 1 State equation parameters

	Euro Area	United States		
	(OECD)	(OECD)	(CBO)	(CBO)
	1976–1998	1976–1998	1976–1998	1960–1998
δ	1.07	1.03	0.54	0.64
$\overline{\rho}$	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
$\overline{\alpha}$	0.34	0.39	0.31	0.31
ζ	2.00	1.90	3.10	2.90
c	0.1	0.1	0.1	0.1
σ_ρ	0.20	0.20	0.20	0.20
σ_α	0.10	0.10	0.10	0.10
σ_e	0.96	0.85	0.89	1.06

Table 2 Minimax solution

Bounds				$F_p(x, v)$	$F_z(x, v)$
$\frac{1}{2}\sigma_u$	$\frac{1}{2}\sigma_e$	$\frac{1}{2}\sigma_\alpha$	$\frac{1}{2}\sigma_\rho$	48.63	80.63
σ_u	σ_e	σ_α	σ_ρ	169.129	224.562
$2\sigma_u$	$2\sigma_e$	$2\sigma_\alpha$	$2\sigma_\rho$	1208.77	1340.59

Table 3 Expected value solution

Uncertainty				$F_p(x, v)$	$F_z(x, v)$
$\mathcal{N}(0, \frac{1}{2}\sigma_u)$	$\mathcal{N}(0, \frac{1}{2}\sigma_e)$	$\mathcal{N}(\overline{\alpha}, \frac{1}{2}\sigma_\alpha)$	$\mathcal{N}(\overline{\rho}, \frac{1}{2}\sigma_\rho)$	15.67	30.13
$\mathcal{N}(0, \sigma_u)$	$\mathcal{N}(0, \sigma_e)$	$\mathcal{N}(\overline{\alpha}, \sigma_\alpha)$	$\mathcal{N}(\overline{\rho}, \sigma_\rho)$	37.76	60.91
$\mathcal{N}(0, 2\sigma_u)$	$\mathcal{N}(0, 2\sigma_e)$	$\mathcal{N}(\overline{\alpha}, 2\sigma_\alpha)$	$\mathcal{N}(\overline{\rho}, 2\sigma_\rho)$	137.16	179.96

size of the bounds. In Table 3 results of minimizing the expected value are presented. It can be observed that the optimal objective function of the expected value model also increases as the variance of the random variables increase. However, the increase is not as much as in the minimax model. In Figs. 4 and 5 we show the results of linear and exponential fits of how the objective function value changes as Δ changes. It is clear from Fig. 4 that, in the minimax model, the relationship between the size of the bounds and the optimal objective function value is exponential. The linear fit is not at all an accurate representation of this relationship. Similarly comments apply for the expected value model. From Fig. 5 it can be argued that the linear fit is more accurate than in the minimax model. However, the exponential fit is clearly more precise for both models.

It can be observed from the results (compare Tables 2 and 3) that expectation of the loss is always lower than the worst–case. Results of cross evaluation are presented

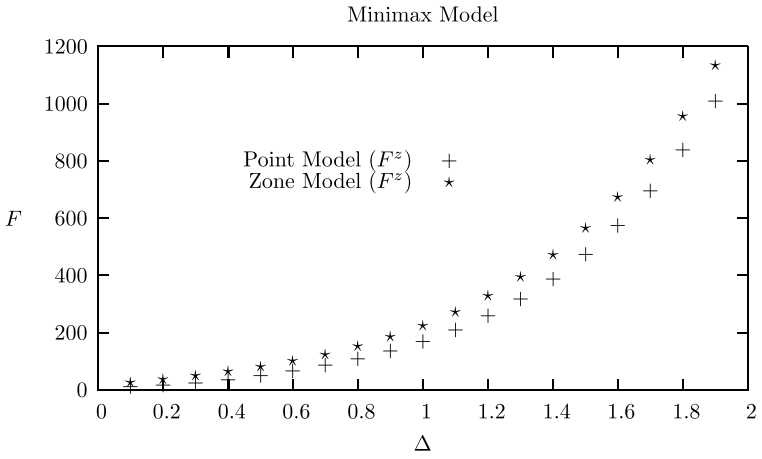

Fig. 2 Objective function increase with size of bounds

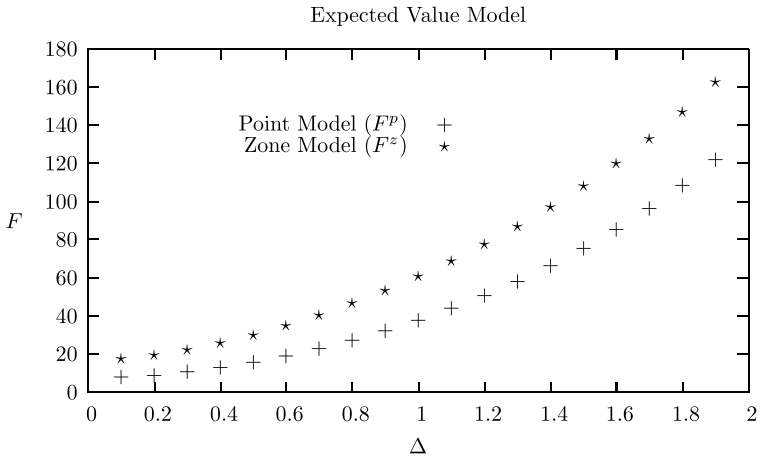

Fig. 3 Objective function increase with size of bounds

in Tables 4 and 5. In Table 4 the second column presents the worst case optimum for the two models. In the last column we show the result of the Monte Carlo simulation if the minimax strategy is followed. Similarly for Table 5, the second column shows the expected value solution for the two models. In the last column of Table 5 we present the deterioration of the optimal objective function value if the expected value optimal solution is implemented but the worst case is realized. In order to understand this relationship better, we plot the value of the objective function for each time period separately in Figs. 6 and 7. In Fig. 6 we show how the objective function will deteriorate if the worst case happens but the expected value policy is implemented. It can be seen that the worst case strategy is always better, even in a time period by time period comparison, up to period 30. Then the expected value strategy seems to

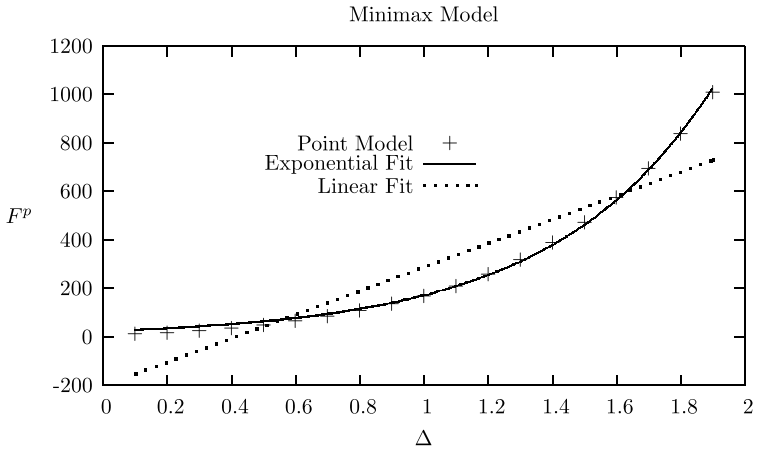

Fig. 4 Optimal objective function value, linear & exponential fit with bounds

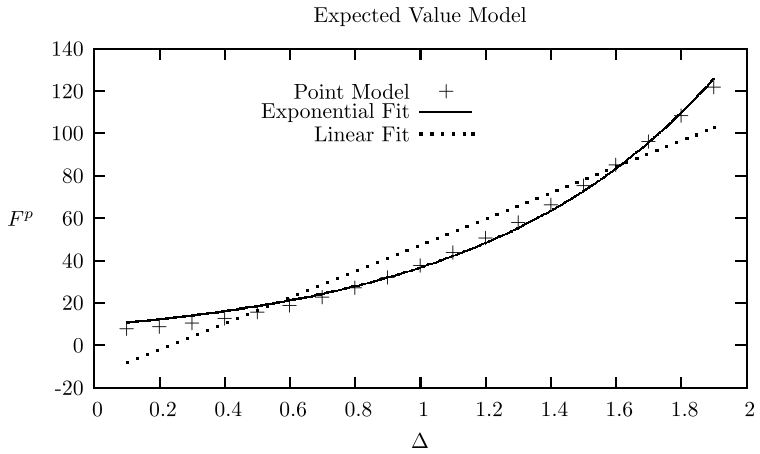

Fig. 5 Optimal objective function value, linear & exponential fit with variance

be better, in the end the two converge. This result is due to the fact that we use a discount factor ($\beta^t = 0.9^t$) in our model. It is also an indication that the optimal solution will greatly improve if we used a dynamic feedback rule. We note that the minimax strategy always outperforms the expected value strategy, if the worst case happens, provided that we add all the time periods together. Note that as Δ decreases from 4.5 in Fig. 6 to 1.0 in Fig. 7 the two strategies start to produce similar results.

In Fig. 8 we show how the objective function deteriorates as the bound of the uncertainties increase. Risk in Fig. 8 is defined as:

$$\text{Risk} = F^p(\overline{x}, v^{wc}) - F^p(x^{wc}, v^{wc}),$$

Table 4 Cross evaluation–minimax solution

Minimax optimum					
Bounds				Worst–case (F_p, F_z)	Exp. val. (F_p, F_z)
$\frac{1}{2}\sigma_u$	$\frac{1}{2}\sigma_e$	$\frac{1}{2}\sigma_\alpha$	$\frac{1}{2}\sigma_\rho$	(48.63, 80.63)	(18.58, 47.53)
σ_u	σ_e	σ_α	σ_ρ	(169.129, 224.562)	(48.30, 75.43)
$2\sigma_u$	$2\sigma_e$	$2\sigma_\alpha$	$2\sigma_\rho$	(1208.77, 1340.59)	(139.88, 183.07)

Table 5 Cross evaluation–expected value solution

Expected value optimum					
Bounds				Exp. val. (F_p, F_z)	Worst–case (F_p, F_z)
$\mathcal{N}(0, \frac{1}{2}\sigma_u)$	$\mathcal{N}(0, \frac{1}{2}\sigma_e)$	$\mathcal{N}(\overline{\alpha}, \frac{1}{2}\sigma_\alpha)$	$\mathcal{N}(\overline{\rho}, \frac{1}{2}\sigma_\rho)$	(15.67, 30.13)	(57.65, 87.52)
$\mathcal{N}(0, \sigma_u)$	$\mathcal{N}(0, \sigma_e)$	$\mathcal{N}(\overline{\alpha}, \sigma_\alpha)$	$\mathcal{N}(\overline{\rho}, \sigma_\rho)$	(37.56, 60.91)	(186.74, 248.18)
$\mathcal{N}(0, 2\sigma_u)$	$\mathcal{N}(0, 2\sigma_e)$	$\mathcal{N}(\overline{\alpha}, 2\sigma_\alpha)$	$\mathcal{N}(\overline{\rho}, 2\sigma_\rho)$	(137.16, 179.96)	(1223.46, 1356.01)

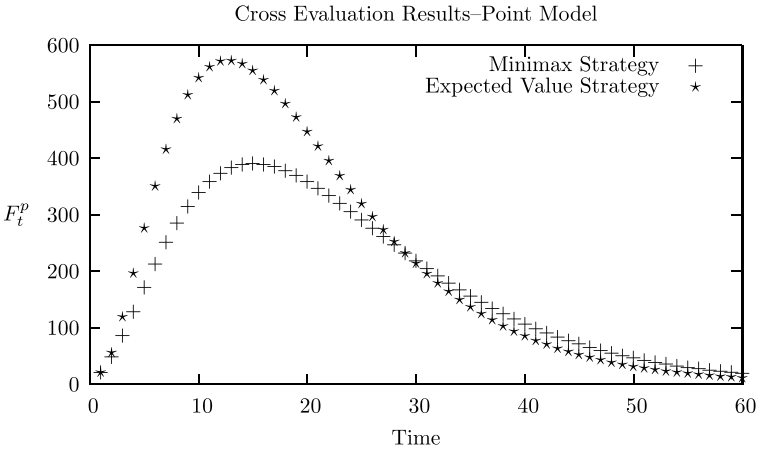

Fig. 6 Behavior of optimal strategies given that the worst case realizes, $\Delta = 3.5$

where x^{wc}, $\overline{x}$ and v^{wc} represent the minimax strategy, expected value strategy and worst case realization of the uncertainties respectively. Again it can be seen that as the magnitude of the uncertainties increase the "risk" increases exponentially. It is also of interest to assess what happens, on average, when the worst case strategy is implemented. Figures 9 and 10 show that, on average, the expected value strategy is better than the worst case strategy independently of the time period. Again it is interesting to study this relationship as the size of the uncertainties increase. In Fig. 11 we display results concerning the "cost" of implementing the worst–case strategy.

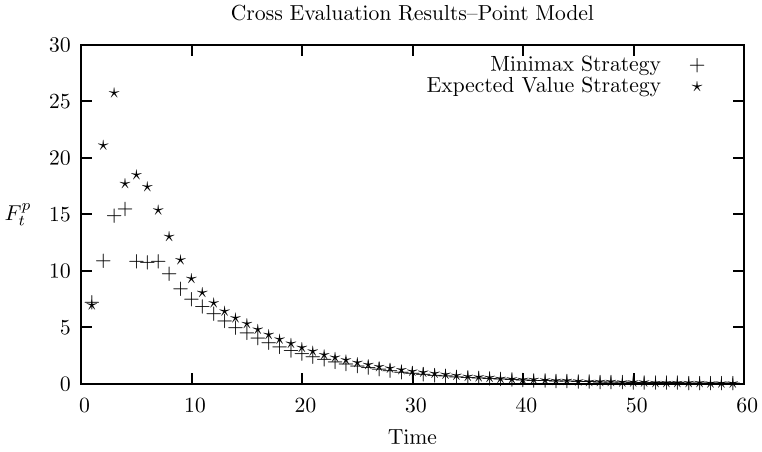

Fig. 7 Behavior of optimal strategies given that the worst case realizes, $\Delta = 1.0$

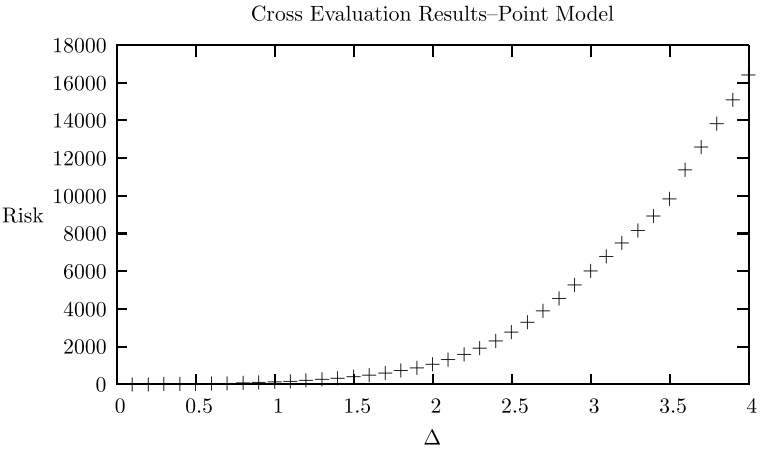

Fig. 8 Deterioration of optimal objective function if worst case happens but expected value strategy is implemented

Cost in Fig. 11 is defined as:

$$\text{Cost} = EF(x^{wc}, v) - EF(\overline{x}, v),$$

where x^{wc}, and $\overline{x}$ represent the minimax strategy and expected value strategy respectively. Again it can be seen that as the magnitude of the uncertainties increase the "cost" increases exponentially. Fig. 8 suggests that the risk of not implementing the minimax strategy increases exponentially as the size of the uncertainty increases. Fig. 11 suggests that the cost of implementing the minimax strategy also raises exponentially as the size of the uncertainties increase. However, the two increases are on different scales, and for Δ sufficiently small ($\Delta < 3$) the cost of the minimax strategy is low, compared to the cost of not implementing the robust strategy. In conclusion,

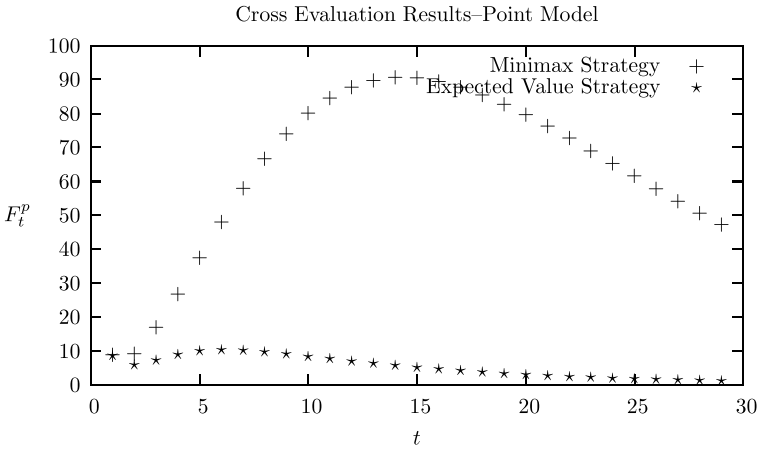

Fig. 9 Expected value of the worst case strategy compared with the optimal expected value strategy, $\Delta = 3.5$

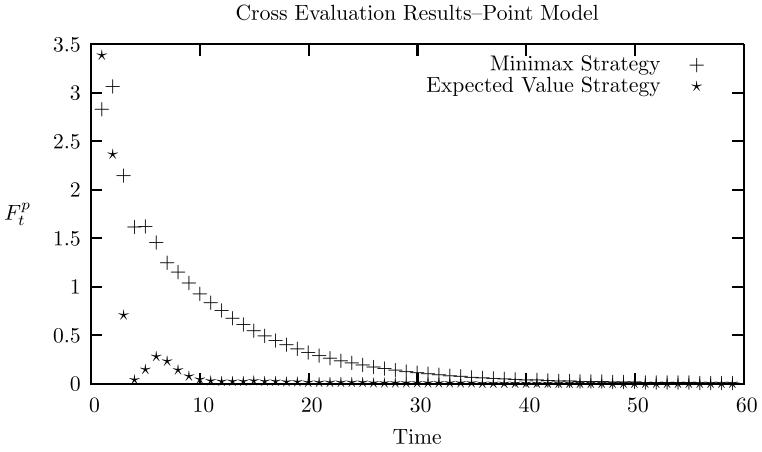

Fig. 10 Expected value of the worst case strategy compared with the optimal expected value strategy, $\Delta = 3.5$

robustness in this model is not as expensive as one might think. If Δ is large, then one would not expect either strategy to perform well.

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. Our numerical experiments also indicate the importance of the size of the uncertainties. As the size of the uncertainties increase, then the expected value strategy can lead to an exponential increase in the objective function if the worst case happens. This can be viewed as a risk taken by the decision maker by implementing a strategy that is not robust against the worst case.

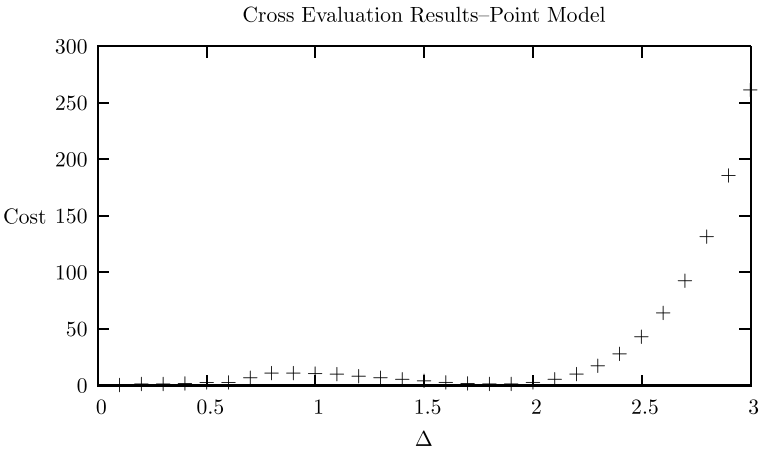

Fig. 11 Average optimal objective function value difference between expected value strategy and minimax strategy as size of uncertainty increases

On the other hand if the worst case strategy is implemented, then the cost of robustness increases exponentially with the size of the uncertainties. These results stress the importance of considering both decision models.

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 [14] was used. The results presented in Sect. 4 showed that, although the expected value optimization performed better on average, the worst–case optimal strategy provided robust solutions, that performed better under the worst–case scenarios. Cross evaluation of worst–case scenarios for expected value strategy indicates that performance deterioration for the latter could be a serious issue.

Acknowledgements The authors thank two anonymous referees for their fruitful suggestions that helped improve the paper. This research was supported by two EPSRC grants—GR/T02560/01 and EP/C513584/1.

Appendix: 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
$x_{l+1} = \arg\min_x \mathrm{Var}_v(F(x, v))$ (from (18))

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

References

1. Başar, T., Bernhard, P.: H^∞-Optimal Control and Related Minimax Design Problems, 2nd edn. Systems & Control: Foundations & Applications. Birkhäuser Boston, Boston (1995). A dynamic game approach
2. Birge, J.R., Wets, R.J.-B.: Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse. Math. Program. Stud. **27**, 54–102 (1986). Stochastic programming 84. I
3. Chow, G., Corsi, P.: Evaluating the Reliability of Macroeconomic Models. Wiley, New York (1982)
4. Dudley, R.M.: Real analysis and probability, Cambridge Studies in Advanced Mathematics, vol. 74. Cambridge University Press, Cambridge (2002)
5. Ermoliev, Y., Gaivoronski, A., Nedeva, C.: Stochastic optimization problems with incomplete information on distribution functions. SIAM J. Control Optim. **23**(5), 697–716 (1985)
6. Fair, R.C.: Specification Estimation and Analysis of Macroeconometric Models. Harvard University Press, Cambridge (1984)
7. Hall, S.G., Henry, S.G.B.: Macroeconomic Modelling. North-Holland, Amsterdam (1988)
8. Hall, S.G., Stephenson, S.J.: Optimal control of stochastic non-linear models. In: Christodoulakis, N. (ed.) Dynamic Modelling & Control of National Economies. Pergamon, Elmsford (1990)
9. Hansen, L., Sargent, T.J., Tallarini, T.D.: Robust permanent income and pricing. Rev. Econ. Stud. **66**, 873–907 (1999)
10. Hansen, L.P., Sargent, T.J.: Robust estimation and control under commitment. J. Econ. Theor. **124**(2), 258–301 (2005)
11. Jacobson, D.H.: New interpretations and justifications for worst case min–max design of linear control systems. In: Guardabassi, G., Locatelli, A., Rinaldi, S. (eds.) Sensitivity, Adaptivity and Optimality: Proceedings of the Third IFAC Symposium, June 1973
12. Kall, P.: Approximation to optimization problems: an elementary review. Math. Oper. Res. **11**(1), 9–18 (1986)
13. Kolbin, V.V.: Decision Making and Programming. World Scientific, River Edge (2003). Translated from the Russian by V.M. Donets
14. Orphanides, A., Wieland, V.: Inflation zone targeting. Eur. Econ. Rev. **44**(7), 1351–1387 (2000)
15. Rustem, B.: Stochastic and robust control of nonlinear economic systems. Eur. J. Oper. Res. **73**, 304–318 (1994)
16. Rustem, B.: Algorithms for Nonlinear Programming and Multiple-Objective Decisions, Wiley-Interscience Series in Systems and Optimization. Wiley, Chichester (1998)
17. Rustem, B., Howe, M.: Algorithms for Worst–Case Design and Applications to Risk Management. Princeton University Press, Princeton (2002)
18. Rustem, B., Becker, R.G., Marty, W.: Robust min-max portfolio strategies for rival forecast and risk scenarios. J. Econ. Dyn. Cont. **24**(11–12), 1591–1621 (2000)
19. Rustem, B., Zakovic, S., Wieland, V.: A continuous minimax problem and its application to inflation targeting. In: Zaccour, G. (ed.) Decision and Control in Management Science. Kluwer Academic, Dordrecht (2002)

20. Rustem, B., Zakovic, S., Parpas, P.: Convergence of an interior point algorithm for continuous minimax. J. Optim. Theory Appl. (2008, to appear)
21. Tetlow, R.J., Muehlen, P.: Robust monetary policy with misspecified models: does model uncertainty always call for attenuated policy? J. Econ. Dyn. Control **25**, 911–949 (2001)
22. Zakovic, S., Pantelides, C., Rustem, B.: An interior point algorithm for computing saddle points of constrained continuous minimax. Ann. Oper. Res. **99**, 59–77 (2001)