CENTRO STUDI IN ECONOMIA E FINANZA

CENTRE FOR STUDIES IN ECONOMICS AND FINANCE

WORKING PAPER NO. 68

On the Relation between Robust and Bayesian Decision Making

Klaus Adam

September 2001

DIPARTIMENTO DI SCIENZE ECONOMICHE - UNIVERSITÀ DEGLI STUDI DI SALERNO
Via Ponte Don Melillo - 84084 FISCIANO (SA)
Tel. 089-96 3167/3168 - Fax 089-96 3169 – e-mail: csef@unisa.it

CSEF WORKING PAPER NO. 68

On the Relation between Robust and Bayesian Decision Making

Klaus Adam*

Abstract

This paper compares Bayesian decision theory with robust decision theory where the decision maker optimizes with respect to the worst state realization. For a class of robust decision problems there exists a sequence of Bayesian decision problems whose solution converges towards the robust solution. It is shown that the limiting Bayesian problem displays infinite risk aversion and that its solution is insensitive (robust) to the precise assignment of prior probabilities. Moreover, the limiting Bayesian objective turns out not to be time separable even if the objective function of the robust decision makers displays time separability.

* Fondazione CARISAL and CSEF, University of Salerno. E-mail address: klaus.adam@iue.it

Table of contents

1 Introduction

In recent years robust or maxmin decision theory has been put forward as an alternative to standard Bayesian decision theory in macroeconomics (e.g. Hansen and Sargent (2000), (2001)). The key idea behind robust decision theory is that agents might face uncertainty that they cannot quantify in terms of prior probabilities because 'too little is known' to do so. By introducing uncertainty aversion robust decision theory provides a mean to calculate optimal decisions in the absence of prior probabilities, see Gilboa and Schmeidler (1989) for an axiomatization.[1]

A key motivation for introducing robust decision makers into macroeconomic models is that such models can explain behavior that seems not to be rational from a Bayesian perspective and thereby improve the descriptive performance of otherwise standard macroeconomic models. Hansen et al. (1999), for example, show that a slight preference for robustness can explain a substantial part of the observed equity premiums.

Despite its increasing popularity in applied macroeconomics (e.g. Onatski and Stock (2000), Tetlow and von zur Muehlen (2001)), the relation of robust decision theory to standard Bayesian decision theory has received little attention. Yet, it is important to understand the links between the two problems since these might inform us in which ways robust decision makers may alter and improve the descriptive performance of macroeconomic models.

The present paper tries to fill some of this gap and shows that robust decision problems can be interpreted in terms of the limit of a sequence of Bayesian decision problems.

Considering a simple class of robust decision problems, I show that there is a sequence of Bayesian decision problems with infinitely increasing risk aversion that has the property that the associated optimal decisions converge to the optimal robust decision. Convergence is robust to the precise assignment of prior probabilities by the Bayesian as long as strictly positive probability is assigned to all states over which the robust decision faces unquantifiable uncertainty.

[1] I use the term robust decision theory synonymous to the term maxmin decision theory, as put forward by Gilboa and Schmeidler (1989). Hansen et. al. (2001) have shown how these two classes of problems can be linked.

The independence of the convergence result from the precise assignment of prior probabilities delivers a Bayesian interpretation of robust decision theory: it represents the choice of a particular objective function that has the property that optimal *Bayesian* decisions are insensitive (or robust) to many different priors. This may be seen as a reply to Sims (2001) who criticized the use of minimax approaches in policy making.

However, the interpretation in terms of a Bayesian setting is not fully satisfying and some problems remain that cannot be reconciled with Bayesian decision theory: since the additional maximization operator which appears in robust decision problems induces infinite risk aversion only over the domain of maximization, the decision maker's risk aversion outside this domain remains unaltered.

Therefore, robust decision makers seem to have a split personality from the viewpoint of a Bayesian and act as if they faced two degrees of risk aversion: the usual one with respect to uncertainty to which they can assign prior probabilities and an infinite one with respect to unquantifiable uncertainty.

With regard to macroeconomic applications, a Bayesian interpretation of robust decisions in terms of extreme degrees of risk-aversion may partly explain why robust decision theory has the potential to generate larger risk premia, see Hansen et al. (1999).

Besides ever increasing risk-aversion, the sequence of Bayesian decision problems has a second interesting property: in general utility is not time separable even if the objective function of the robust decision maker displays time separability. This result is due to the fact that the worst case evaluated by the robust decision maker depends on full decision vector and not simply on the decision of a single period. Like this the additional maximization operator in the robust decision problem induces non separability of the ratio of marginal utilities in the equivalent Bayesian problem.

This observation suggests that macroeconomic models with robust decision makers should deliver results similar to models with Bayesian decision makers with non-separable utility. Since habit persistence in consumption services may increase the market price of risk in a model of Bayesian decision making, it seems natural to expect a similar increase in this price in models with robust decision makers but seemingly time separable period loss functions, as found by Hansen et al. (1999).

The next section introduces the decision problem and describes the robust and Bayesian approach to its solution. Section 3 derives the convergence result which is illustrated in section 4 with the help of a simple example. Finally, section 5 extends the setup to infinite dimensional decision problems with discounting.

2 Bayesian and Robust Decision Problems

Consider a decision maker whose objective can be described by a simple loss function which depends on a decision vector $x \in R^n$ and an unknown state of the world s:

$$L(x, s)$$

$L(\cdot, s)$ is assumed to be twice continuously differentiable and strictly convex for all s. The state of the world s is assumed to belong to some finite and known set $\Omega_s = \{s_1, \dots s_I\}$ and the decision must be chosen from a compact set $\Omega_x \subset R^n$ of feasible decisions.

First, consider a Bayesian decision maker. Based on Savage's axioms such a decision maker can construct subjective prior probabilities p_i $(i = 1, \dots n)$ that describe the likelihood with which the decision maker believes that state s_i will realize. The Bayesian then acts to

$$\min_{x \in \Omega_x} E\left[L(x, s)\right] = \min_{x \in \Omega_x} \sum_{i=1}^{I} L(x, s_i) p_i \tag{1}$$

Next, consider what has been called a robust decision maker who cannot assign meaningful priors to the realization of the state s. One reason might be that some of Savage's axioms do not hold, e.g. if there is no random variable with uniform distribution that allows for the calibration of probabilities.

Uncertainty that cannot be quantified in terms of subjective probabilities has been called Knightian uncertainty in the literature. The existence of Knightian uncertainty opens many possible ways for modeling the decision problem. One possible and intuitive way, suggested by Blinder (1998), is to simply average over the states of the world. The resulting decision problem would be equivalent to a Bayesian decision problem with $p_i = \frac{1}{I}$ $(i = 1, \dots, I)$.

Yet, the most widely used method to model robust decisions and the one I will consider in this paper advocates to let the decision maker choose an action x that minimizes the maximum possible loss associated with action x. In mathematical terms

$$\min_{x \in \Omega_x} \max_{s \in \Omega_s} L(x, s) \tag{2}$$

Let x_r^* denote the solution to minimization part of this problem.

It is useful to rewrite this optimization problem as follows

$$\min_{x \in \Omega_x} R(x) \qquad \text{with}$$
$$R(x) \equiv \sum_{i=1}^{I} L(x, s_i) I(x, s_i) \tag{3}$$

where $I(x, s_i)$ is an indicator function that is one if s_i is a maximizer of $L(x, s_i)$ and that is zero otherwise.[2]

Rewriting the robust objective in this form helps to highlight the relation to the Bayesian problem (1) since the indicator functions in (3) look almost like prior probabilities. The difference between the Bayesian prior and the robust 'prior' is that the robust decision maker always puts all probability weight on the worst state associated with a given decision x; since this worst state may shift with x, the 'prior' of the robust decision maker may shift with the chosen decision, which is not the case for a Bayesian decision maker.

3 Linking Bayesian and Robust Decision Problems

The objective of this section is to establish a link between the two decision problems. The main idea is to manipulate the objective function of the Bayesian decision problem in a way that the Bayesian's objective function will have the same minimum as the robust objective.

Since the loss function depends on the action x,altering the Bayesian's loss function is a back-door through which one can cause the Bayesian to behave as

[2] If there are several maximizers I define the indicator function to be 1 only for the state with the lowest index i.

if her priors were changing across actions. In particular, if the Bayesian was to maximize a transformed loss function $T(L(x,s))$ with the property that

$$T(L(x,s)) = L(x,s) \cdot \frac{I(x,s)}{p_s} \tag{4}$$

where p_s is the prior probability for state s, then the Bayesian problem would be identical to the robust decision problem:

$$\begin{aligned}
&\min_{x \in \Omega_x} E\left[T(L(x,s))\right] \\
&= \min_{x \in \Omega_x} \sum_{i=1}^{I} L(x,s_i) \frac{I(x,s_i)}{p_i} p_i \\
&= \min_{x \in \Omega_x} R(x)
\end{aligned}$$

Of course, such a transformed 'loss function' is not a loss function in the strict sense since it depends on prior probabilities.

Given that direct equivalence between the two problems requires a Bayesian loss that depends on priors, the strategy to is to construct a sequence of transformed loss functions $T^k(L(x,s))$ for the Bayesian problem with the property that these transformed loss functions are *independent* from the prior. At the same time the solution x_k^* to

$$\min_{x \in \Omega_x} E\left[T^k(L(x,s))\right] \tag{5}$$

should converge to the robust solution x_r^k as k increases without bound, i.e.

$$\lim_{k \to \infty} \|x_k^* - x_r^*\| = 0$$

I define the sequence of transforming functions $T^k(\cdot)$ as[3]

$$T^k(L) = e^{kL}$$

Since $T^k(\cdot)$ is increasingly convex a Bayesian with objective $T^k(L)$ will become increasingly risk averse in terms of the coefficient of absolute risk aversion. As a result, the value of the transformed loss $T^k(L)$ increases disproportionately with the size of the loss L. Intuitively, this implies that the largest of all losses

[3]The particular sequence T^k is just chosen for convenience and other sequences might give the same result.

L associated with some action x obtains increasing relative weight. This should move the solution to the Bayesian decision problem closer and closer to the robust solution.

Theorem 1 below shows that this is the case. Robust decisions can be interpreted as decisions of a Bayesian with an infinite degree of risk-aversion and arbitrary strictly positive priors over the domain to which the robust decision maker cannot assign prior probabilities. In Bayesian terms robustness represents a choice of a particular objective function, which has the property that optimal decisions are robust to the assignment of prior probabilities.

Theorem 1 *Let x_k^* denote the solution to the transformed Bayesian decision problem (5) with prior probabilities $p_i > 0$ $(i = 1, \dots n)$. Let x_r^* denote the solution to the robust decision problem (3). Then*

$$\lim_{k \to \infty} \|x_k^* - x_r^*\| = 0$$

Proof of theorem 1: Rename states s such that at x_r^*

$$L(x_r^*, s_1) \geq L(x_r^*, s_2) \geq \dots \geq L(x_r^*, s_I)$$

and let

$$\Omega_{\max} = \{i | L(x_r^*, s_i) = L(x_r^*, s_1)\}$$

Then I first show the following auxiliary result:

Lemma 2 $\forall\, \delta > 0$ *sufficiently small* $\forall\ d \in R^n$ *with* $\|d\| = \delta\ \exists\ \varepsilon > 0$ *independent of d and a state $i \in \Omega_{\max}$ s.t.*

$$L(x_r^* + d, s_i) - L(x_r^*, s_i) > \varepsilon$$

Proof of lemma 2:. The difference can be expressed as

$$L(x_r^* + d, s_i) - L(x_r^*, s_i) = \nabla L(x_r^*, s_i)d + d'\nabla^2 L(x_r^*, s_i)d + O(3) \qquad (6)$$

where $O(3)$ is a third order approximation error. Consider the first order term: From the optimality of x_r^* follows that

$$\nabla\, L(x_r^*, s_i)d \geq 0 \qquad (7)$$

for some $i \in \Omega_{\max}$. Next, fix such an i and consider the second order term. Since $\nabla^2 L(x_r^*, s_i)$ is normal and positive definite, we have

$$\nabla^2 L(x_r^*, s_i) = U_i' D_i U_i$$

where U_i is unitary and

$$D_i = diag(\lambda_{i,1} \dots \lambda_{i,n})$$

with $\lambda_{i,j} > 0$ being the eigenvalues of $\nabla^2 L(x_r^*, s_i)$. Then defining $\lambda_{i,\min} = \min_j \lambda_{i,j}$

$$\begin{aligned} d'\nabla^2 L(x_r^*, s)d &= d'U_i' D_i U_i d \\ &\geq \lambda_{i,\min} d'U_i' U_i d \qquad (8) \\ &= \lambda_{i,\min} d'd \qquad (9) \\ &= \lambda_{i,\min} \delta^2 \qquad (10) \end{aligned}$$

Letting $\lambda_{\min} = \min_{i \in \Omega_{\max}} \lambda_{i,\min}$ it follows from (6), (7), and (10) that

$$L(x_r^* + d, s_i) - L(x_r^*, s_i) \geq \lambda_{\min}\delta^2 + O(3)$$

Choosing δ sufficiently small the third order approximation error can be made arbitrarily small, e.g. smaller than $\frac{\lambda_{\min}\delta}{2}$, then choosing $\varepsilon = \frac{\lambda_{\min}\delta}{2}$ establishes the claim.■

Next, normalize the transformed objective of the Bayesian decision maker (5) as follows

$$\overline{L}^k(x) = \frac{1}{e^{kL(x_r^*, s_1)}} \sum_{i=1}^{I} e^{kL(x, s_i)} p_i \qquad (11)$$

Maximizing (11) delivers the same solution as maximizing (5). The limit of $\overline{L}^k(x_r^*)$ for $k \to \infty$ exists and is given by:

$$\lim_{k\to\infty} \overline{L}^k(x_r^*) = \sum_{i \in \Omega_{\max}} p_i$$

Next, consider $\overline{L}^k(x_r^* + d)$ with $d \in R^n$ and $\|d\| = \delta$, δ sufficiently small. From the auxiliary lemma and (11) it follows that

$$\overline{L}^k(x_r^* + d) > \frac{e^{k(L(x_r^*, s_1) + \varepsilon)}}{e^{kL(x_r^*, s_1)}} p_{\min}$$

where $p_{\min} = \min_i p_i$. Therefore, there exists a $\overline{k} < \infty$ such that for all $k > \overline{k}$

$$\overline{L}^k(x_r^* + d) > \overline{L}^k(x_r^*)$$

>From the strict convexity of $\overline{L}^k(\cdot)$ it follows that the minimum x_k^* of $\overline{L}^k(\cdot)$ must be within distance δ from x_r^* for all $k > \overline{k}$, which establishes the claim. ■

4 An Example

Consider the following simple loss function, which has been considered amongst others by Brainard (1967) and Onatski (2000)

$$L(x, s) = (sx - \pi^*)^2$$

where the variable π^* denotes the inflation target and where sx is the inflation rate that results when the decision maker chooses policy x and the state of the world is given by s. If x is the real interest rate, then the factor s represents the sensitivity of the economy's inflation rate to the real interest rate, a number likely to be unknown to the policy maker. Moreover, the policy maker might have unquantifiable uncertainty over the various values of s.

To keep things simple suppose that the desired target inflation rate is $\pi^* = 2$ and that there are only two potential multipliers $s_l < s_h$ with $s_l = 1$ and $s_h = 3$.

The loss functions associated for each of these multipliers are shown in figure 1. The dotted line in the graph indicates the maximum loss associated with each action. As figure 1 shows, the optimal robust decision that minimizes the maximum loss is given by $x_r^* = 1$.[4]

Next, consider a Bayesian central bank and let its subjective prior place equal probability on each of the two multipliers. The optimal Bayesian decision is easily calculated to be $x^* = 0.8$. This is a less aggressive policy than the one chosen by the robust decision maker. The reason for this result is that the Bayesian trades off the gains and losses across the different realizations of s. Clearly at $x = 1$ the loss functions in figure 1 have different absolute slope coefficients depending on whether $s = s_l$ or $s = s_h$. Therefore, the Bayesian

[4] Since there is no uncertainty about the sign of the parameters a the optimal robust decision coincides with the optimal decision under certainty equivalence, as noted by Onatski (2000).

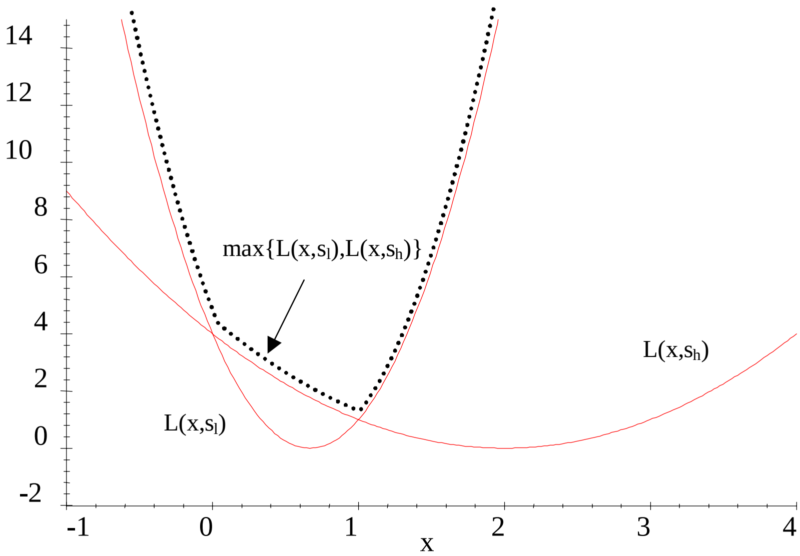

Figure 1: Loss functions

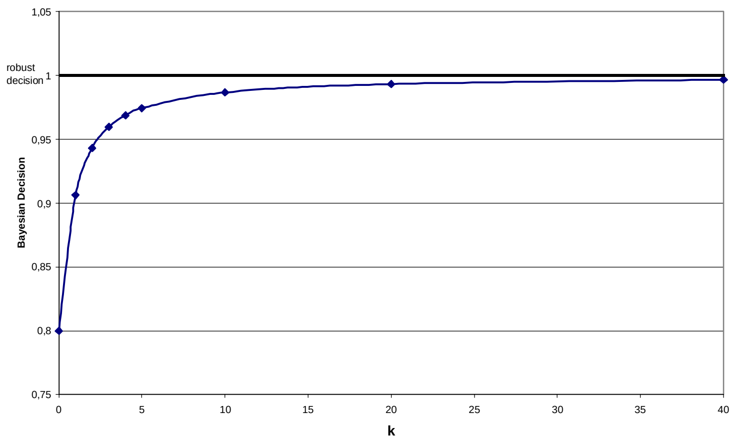

Figure 2: Optimal Bayesian Decision as a Function of k

has an incentive to decrease the interest rate below 1 since the gains made for the realization s_h will exceed the losses for realizations s_l, given the prior probabilities assigned to these states.

When the Bayesian's objective function is subjected to increasingly convex transformations $T^k(\cdot)$, then she becomes increasingly risk averse. This implies that the gains for the state s_h will be appreciated less relative to the potential increase in the loss for the state s_l. Graphically one can interpret this as figure 1 being scaled in the direction of the y-axis with each point being scaled by a factor that is increasing with its distance from the x-axis. As a result, the slope of $L(x, s_l)$ to the left of $x = 1$ increases much faster than the absolute value of the slope of $L(x, s_h)$ to the left of this point and the Bayesian decision will approach the robust decision. Figure 2 shows the optimal Bayesian decision as a function of k and illustrates that the robust decision is approached rather rapidly as k increases.[5]

5 Extensions

The loss function considered so far was rather general but has assumed a finite dimensional decision vector. Yet, macroeconomists tend to use infinite horizon models with infinite dimensional decision vectors. This section shows that the results of the previous section extend in a natural way to the infinite horizon problems with discounting as they are typically used in macroeconomics.

Consider the following loss function

$$L(x, s) = \sum_{t=0}^{\infty} \beta^t l(x_t, s)$$

where $x_t \in R^n$ is the period t decision, the vector $x = (x'_0, x'_1, \dots)'$ the stacked period decisions, and $\beta < 1$ a discount factor. The period loss function $l(\cdot, s)$ is assumed to be strictly convex and twice continuously differentiable for all s. The period decision x_t must be chosen from a compact set of feasible decisions Ω_t that might depend on past decisions. Furthermore, there is a compact set $\Omega_x \subset R^n$ such that $\Omega_t \subset \Omega_x$ for all t.

[5] As the example reveals, the speed of convergence will depend on the prior. If less weight is attached to the worst state of the optimal robust decision then convergence will be slower (see also the last bit in the proof of theorem 1).

The robust decision maker minimizes

$$\min_{\{x_t|x_t\in\Omega_t\}} \max_{s\in\Omega_s} \sum_{t=0}^{\infty} \beta^t l(x_t, s_i) \tag{12}$$

To construct the transformed Bayesian problem it might seem natural to alter directly the period loss function $l(\cdot,\cdot)$ to preserve the time separability of the objective, e.g. to let the Bayesian minimize

$$\min_{\{x_t|x_t\in\Omega_t\}} \sum_{i=1}^{I} \sum_{t=0}^{\infty} \beta^t e^{kl(x_t,s_i)} p_i \tag{13}$$

However, the solution to this problem will *not* necessarily converge to the solution of the robust decision problem, as the following example shows.

Example 3 *Let the optimal robust decision be given by* $x_r^{*\prime} = (x_{r,0}^{*\prime}, x_{r,1}^{*\prime}, \ldots)$ *and the state* s_i $(i = 1, 2)$ *that maximizes the loss for this and any neighboring decisions is given by* s_1*. Next consider the decision*

$$x' = x_r'^{*} + (d', 0, 0, 0\ldots)$$

which is equal to x_r^* *except for the first period. Suppose that altering the decision from* x_r^* *to* x *causes the loss in period zero to increase by* $\gamma_1 > 0$ *units, which makes* x *suboptimal for the robust decision maker.*

Next, consider a Bayesian decision maker with objective (13) who considers a deviation from x_r^* *to* x*. The change* Δ_k *in the first period loss is given by*

$$\Delta_k = (e^{k(l(x_{r,0}^*,s_1)+\gamma_1)} - e^{kl(x_{r,0}^*,s_1)})p_1 + \left(e^{k(l(x_{r,0}^*,s_2)+\gamma_2)} - e^{kl(x_{r,0}^*,s_2)}\right) p_2 \tag{14}$$

where $\gamma_2 = l(x_{r,0}^* + d, s_2) - l(x_{r,0}^*, s_2)$*. Suppose* $\gamma_2 < 0$ *and* $l(x_{r,0}^*, s_2) > l(x_{r,0}^*, s_1) + \gamma_1 > 0$*, which cannot be excluded, then*

$$\lim_{k\to\infty} \Delta_k = -\infty$$

which indicates that a Bayesian with objective function (13) will prefer x *to* x_{r^*} *for all sufficiently large* k*.*

As it turns out define the transformed loss function as

$$T^k(L(x, s)) = e^{k\left(\sum_{t=0}^{\infty} \beta^t l(x_t,s_i)\right)} \tag{15}$$

and let the Bayesian minimize

$$\min_{\{x_t|x_t\in\Omega_t\}} \sum_{i=1}^{I} e^{k\left(\sum_{t=0}^{\infty}\beta^t l(x_t,s_i)\right)} p_i \tag{16}$$

where p_i are prior probabilities, delivers the desired result. Theorem 4 below shows that, as k increases without bound, the Bayesian solution to problem (16) converges to the robust solution in terms of the following vector norm:

$$\|x\|_\beta = \sum_{t=0}^{\infty} \beta^t x_t' x_t$$

Theorem 4 *Let x_k^* denote the solution to the transformed Bayesian decision problem (16) with prior probabilities $p_i > 0$ $(i = 1, \ldots n)$. Let x_r^* denote the solution to the robust decision problem (12). Then*

$$\lim_{k\to\infty} \|x_k^* - x_r^*\|_\beta = 0$$

Proof of theorem 4: Rename states s such that at x_r^*

$$L(x_r^*, s_1) \geq L(x_r^*, s_2) \geq \ldots \geq L(x_r^*, s_I)$$

and let

$$\Omega_{\max} = \{i | L(x_r^*, s_i) = L(x_r^*, s_1)\}$$

Then I first show the following auxiliary result:

Lemma 5 *$\forall\, \delta > 0$ sufficiently small $\forall\ d \in R^n$ with $\|d\|_\beta = \delta\ \exists\ \varepsilon > 0$ independent of d and a state $i \in \Omega_{\max}$ s.t.*

$$L(x_r^* + d, s_i) - L(x_r^*, s_i) > \varepsilon$$

Proof of lemma 5:. The difference can be expressed as

$$\begin{aligned} L(x_r^* + d, s_i) - L(x_r^*, s_i) &= \nabla L(x_r^*, s_i)d + d'\nabla^2 L(x_r^*, s_i)d + O(3) \\ &= \nabla L(x_r^*, s_i)d + \sum_{t=0}^{\infty} \beta^t d_t' \nabla^2 l(x_{r,t}^*, s_i) d_t + O(3) \end{aligned} \tag{17}$$

where $O(3)$ is a third order approximation error. Consider the first order term: From the optimality of x_r^* follows that

$$\nabla L(x_r^*, s_i)d \geq 0 \tag{18}$$

for some $\widetilde{i} \in \Omega_{\max}$. Next, consider the second order term. Since $\nabla^2 l(x_{r,t}^*, s_i)$ is normal and positive definite

$$\nabla^2 l(x_{r,t}^*, s_i) = U_{i,t}' D_{i,t} U_{i,t}$$

where $U_{i,t}$ is unitary and

$$D_i = diag(\lambda_{i,t,1} \ldots \lambda_{i,t,n})$$

with $\lambda_{i,t,j}$ being the eigenvalues of $\nabla^2 l(x_{r,t}^*, s_i)$. Let $\lambda_{\min}$ denote the minimum eigenvalue of $\nabla^2 l(x_t, s_i)$ over all $x \in \Omega_x$ and $i \in \Omega_s$. Since $\nabla^2 l(\cdot, s_i)$ is continuous, Ω_x is compact, and Ω_s is finite the minimum exists. From the strict convexity of $l(\cdot, s_i)$ follows that $\lambda_{\min} > 0$. Then

$$\begin{aligned} d' \nabla^2 L(x_r^*, s_i)d &= \sum_{t=0}^{\infty} \beta^t d_t' U_{i,t}' D_{i,t} U_{i,t} d_t \\ &\geq \lambda_{\min} \sum_{t=0}^{\infty} \beta^t d_t' U_{i,t}' U_{i,t} d_t \\ &= \lambda_{\min} \sum_{t=0}^{\infty} \beta^t d_t' d_t \\ &= \lambda_{\min} \delta \end{aligned} \tag{19}$$

It follows from (17), (18), and (19) that

$$L(x_r^* + d, s_i) - L(x_r^*, s_i) \geq \lambda_{\min}\delta + O(3)$$

Choosing δ sufficiently small the third order approximation error can be made arbitrarily small, e.g. smaller than $\frac{\lambda_{\min}\delta}{2}$, then choosing $\varepsilon = \frac{\lambda_{\min}\delta}{2}$ establishes the claim.■

Next, normalize the transformed objective of the Bayesian decision maker (15) as follows

$$\overline{L}^k(x) = \frac{1}{e^{k\left(\sum_{t=0}^{\infty} \beta^t l(x_{r,t}^*, s_1)\right)}} \sum_{i=1}^{I} e^{k\left(\sum_{t=0}^{\infty} \beta^t l(x_t, s_i)\right)} p_i \tag{20}$$

Maximizing (20) delivers the same solution as maximizing (15). The limit of $\overline{L}^k(x_r^*)$ for $k \rightarrow \infty$ exists and is given by:

$$\lim_{k\rightarrow\infty} \overline{L}^k(x_r^*) = \sum_{i\in\Omega_{\max}} p_i$$

Next, consider $\overline{L}^k(x_r^* + d)$ with $d \in R^n$ and $\|d\|_\beta = \delta$, δ sufficiently small. From lemma 5 and (20) it follows that

$$\overline{L}^k(x_r^* + d) > \frac{e^{k\left(\varepsilon+\sum_{t=0}^{\infty}\beta^t l(x_t,s_1)\right)}}{e^{k\left(\sum_{t=0}^{\infty}\beta^t l(x_{r,t}^*,s_1)\right)}} p_{\min}$$

where $p_{\min} = \min_i p_i$. Therefore, there exists a $\overline{k} < \infty$ such that for all $k > \overline{k}$

$$\overline{L}^k(x_r^* + d) > \overline{L}^k(x_r^*)$$

>From the strict convexity of $\overline{L}^k(\cdot)$ it follows that the minimum x_k^* of $\overline{L}^k(\cdot)$ must be within distance δ from x_r^* for all $k > \overline{k}$, which establishes the claim. ■

>From theorem 4 follows that the approximating Bayesian loss functions is not time separable, even in the limit. Marginal utility for the transformed Bayesian problem is given by

$$\frac{\partial E\left[T^k(L(x,s))\right]}{\partial x_t} = k\beta^t \sum_i \nabla l(x_t, s_i) e^{k\left(\sum_{h=0}^{\infty}\beta^h l(x_h,s_i)\right)} p_i$$

Therefore, the limit of the ratio of marginal utilities is given by

$$\lim_{k\rightarrow\infty} \frac{\frac{\partial E\left[T^k(L(x,s))\right]}{\partial x_t}}{\frac{\partial E\left[T^k(L(x,s))\right]}{\partial x_{t+j}}}$$

and depends on the states s that maximize $\sum_{h=0}^{\infty} \beta^h l(x_h, s)$, which are a function of the whole decision vector x. Thus, although the loss function of the robust decision maker is separable with respect to the decision vectors x_t, the fact that the robust decision maker evaluates each period decision x_t with respect to the state that generates the greatest *discounted* loss generates non-separability in the equivalent Bayesian problem.

References

Blinder, Alan, *Central Banking in Theory and Practice*, MIT Press, 1998.

Gilboa, Izhak and David Schmeidler, "Macmin Expected Utility Theory with Non-Unique Prior," *Journal of Mathematical Economics*, 1989, *18*, 141–153.

Hansen, Lars P. and Thomas J. Sargent, "Wanting Robustness in Macroeconomics," 2000. Standord University, mimeo.

Hansen, L.P., T.J. Sargent, and T. Tallarini, "Robust Permanent Income and Pricing," *Review of Economic Studies*, 1999, *66*, 873–907.

____, ____, G.A. Tirmuhambetova, and N. Williams, "Robustness and Uncertainty Aversion," 2001. Stanford University Mimeo.

Onatski, Alexei, "Minimax Analysis of Monetary Policy under Model Uncertainty," 2000. Harvard University, Dept. of Economics, mimeo.

____ and James Stock, "Robust Monetary Policy under Model Uncertainty in a Small Model of the U.S. Economy," *NBER Working Paper 7490*, 2000.

Sargent, T.J. and L.P. Hansen, *Robust Control and Filtering for Macroeconomics*, Stanford University book mimeo, 2001.

Sims, Christopher, "Implications of Rational Inattention," 2001. Princeton University Mimeo.

Tetlow, Robert and Peter Von Zur Muehlen, "Robust Monetary Policy with Musspecified Models: Does Model Uncertainty Always Call for Attenuated Policy?," 2001. Federal Reserve Board mimeo.