数理解析研究所講究録
1108巻 1999年 44-61



# NON-HILBERT CONDITIONING AND VALUE OF INFORMATION*

by Hiroyuki Ozaki†

Faculty of Economics
Tohoku University
Kawauchi, Aoba-ku, Sendai 980-8576
JAPAN

First version: January 24, 1995
This version: September 29, 1998

**Abstract**

This paper generalizes the conditional expectation framework by replacing the Hilbert-norm with a more general measure of approximation errors. This enables one to develop the concept of conditioning for non-expectation certainty equivalents. Under this concept a single-agent model, in which the optimal level of information is endogenously determined through the agent's optimization behavior, can be constructed. In this setting it is possible for the anti-preference for information to dominate the benefits of better planning available under more information. In particular, I show that the "biasedness" of this new conditioning is important in endogenizing the choice of information.

*I am grateful to seminar participants at University of Western Ontario, Chubu University and Fukushima University for their helpful comments.

†e-mail address: ozaki@econ.tohoku.ac.jp

## 1. Introduction

This paper generalizes the conditional expectation framework by replacing the Hilbert-norm (*i.e.*, $L^2$-norm) with a more general measure of approximation errors, which enables us to develop the concept of conditioning for non-expectation certainty equivalents, and to construct a single-agent model in which the optimal level of information is endogenously determined through the agent's optimization behavior, and in which the anti-preference for information could dominate the benefits of better planning which would become available under more information.

We first develop the concept of conditioning $M(x|\mathcal{G})$ for non-expectation certainty equivalents. Here $x$ is a random utility and $\mathcal{G}$ is a sub-$\sigma$-algebra which represents partial information held by the agent. To this end, we assume that the agent is endowed with a functional, by which she measures approximation errors. We then define $M(x|\mathcal{G})$ as the best approximation to $x$ within the $\mathcal{G}$-measurable functions when approximation errors are measured by this error functional (its well-definition will be proved in Proposition 2.2). If the error functional is specified by the $L^2$-norm, $M(x|\mathcal{G})$ coincides with the conditional expectation $E(x|\mathcal{G})$. We call the conditioning thus defined as *non-Hilbert conditioning*. Given the error functional, we can *define* the associated certainty equivalent as the best approximation within the constants. Therefore, for any (non-expectation) certainty equivalent $M$, defining its conditioning can be reduced to finding the error functional which defines the given certainty equivalent. The same error functional is now used to define $M(x|\mathcal{G})$ for any $\mathcal{G}$. In this manner, Section 3 develops the conditioning for a broad class of non-expectation certainty equivalents.

While the conditional expectation is unbiased, the non-Hilbert conditioning is typically biased in the sense that $M(M(x|\mathcal{G})) \neq M(x)$. Here we interpret the LHS as the current "mean" utility when the agent expects partial information $\mathcal{G}$ to be obtained in the future, and the RHS as the current "mean" utility when the agent expects no more information. Therefore, it affects agent's welfare whether or not she gets future information. In particular, obtaining some information $\mathcal{G}$ may decrease agent's current utility, which in turn implies that she may prefer not to be informed (rather than be informed) although new information would promote better planning. For example, imagine the person who hates to be informed of the result of a medical examination regardless of a better medical treatment he could get if he found some disease. This aspect is missing in the standard conditional expectation framework. In Section 4, a simple 2-period model will be constructed in which the agent chooses the level of information (together with control variables) in order to maximize her overall utility. We show that it could be the case that being ignorant is preferred to being informed. This new approach would allow us to endogenize the evolution of agent's information in dynamic models as a result of optimization behavior. In Section 5, we briefly mention the future research along this line.

## 2. General Theory

We specify the states of the world by a probability space $(\Omega, \mathcal{F}, \mu)$, where $\mathcal{F}$ is a $\sigma$-algebra of all conceivable events on $\Omega$, and $\mu$ is a probability measure on $\mathcal{F}$. We then define the space of essentially bounded real-valued functions on $\Omega$ by

$$L^\infty(\mathcal{F}) = \{ x : \Omega \to \Re \,|\, x \text{ is } \mathcal{F}\text{-measurable and ess sup}_{\omega\in\Omega} |x(\omega)| < +\infty \}.$$

We identify $x \in L^\infty(\mathcal{F})$ with $y \in L^\infty(\mathcal{F})$ if $x = y$ $\mu$-almost everywhere (a.e.), and simply write as $x = y$. That is, we regard $L^\infty(\mathcal{F})$ as the set of $\mu$-equivalence classes of functions. We denote by $\mathcal{F}^*$ the space of all sub-$\sigma$-algebras of $\mathcal{F}$. When $\mathcal{G} \in \mathcal{F}^*$, we say that $x$ is *almost $\mathcal{G}$-measurable* if there exists $y$ such that $y$ is $\mathcal{G}$-measurable and $x = y$ $\mu$-a.e. We denote by $L^\infty(\mathcal{G})$ the space of all essentially bounded almost $\mathcal{G}$-measurable functions, which becomes a subspace of $L^\infty(\mathcal{F})$. In the current paper, any element of $L^\infty(\mathcal{F})$ or $L^\infty(\mathcal{G})$ may be regarded as a random utility. More specifically, we may consider $x$ as $x(\omega) = u(w(\omega), \omega)$, where $u$ is a state-dependent von Neumann-Morgenstern utility function and $w$ is a random income. We are more concerned with the attitude toward uncertainty (that is, how to aggregate $x$ over states) rather than the attitude toward wealth (that is, the curvature of $u$).

For any $x \in L^\infty(\mathcal{F})$ and for any $\mathcal{G} \in \mathcal{F}^*$, the conditional expectation $E[x|\mathcal{G}]$ is defined by

$$E[x|\mathcal{G}] = \arg\min \{ \| x - z \|_2 \,|\, z \in L^\infty(\mathcal{G}) \}, \tag{1}$$

where $\| \cdot \|_2$ is the $L^2$-norm. This definition is typically applied to $L^2$ spaces, in which case the minimum is always uniquely attained by the orthogonal projection since $L^2(\mathcal{G})$ is a closed subspace of $L^2(\mathcal{F})$. The uniqueness of $E[x|\mathcal{G}]$ is up to a $\mu$-equivalence class and we can always choose a version which is $\mathcal{G}$-measurable. Because $L^\infty \subset L^2$, this definition can be directly applied to the current context. Furthermore, if the above definition is extended to $L^1$ spaces by the standard approximation argument, it coincides with a more common definition of the conditional expectation.[1] Among the implications of this definition are

$$E[x] = E[x|\{\phi, \Omega\}] \tag{2}$$

and

$$E[x] = E[E[x|\mathcal{G}]]. \tag{3}$$

The conditional expectation (1) can be interpreted as the "best" approximation to $x$ within the $\mathcal{G}$-measurable functions when approximation errors are measured by the $L^2$-norm.

[1] That is, $E[x|\mathcal{G}]$ is defined as a $\mathcal{G}$-measurable integrable function which satisfies $(\forall G \in \mathcal{G})$ $\int_G E[x|\mathcal{G}]\, d\mu = \int_G x\, d\mu$. In this regard, see Billingsley (1986, p. 477, 34.15).

We now extend (1) to a more general conditioning concept by replacing the $L^2$-norm in (1) with a more general measure of errors, which we will call error functional. To this end, we first define a partial order on $L^\infty(\mathcal{F})$ by

$$x \preceq y \Leftrightarrow x_+ \leq y_+ \text{ and } x_- \leq y_-,$$

where $x_+$ and $x_-$ are the positive part and the negative part of $x$, respectively. This order measures the "closeness". That is, $z$ is "closer" to $x$ than $z'$ is if $x - z \preceq x - z'$. In this case, *both* of the overestimate and the underestimate (at each state) by $z$ are smaller than those by $z'$. Also note that this order is a proper subset of the order by the absolute values: $x \preceq y \Rightarrow |x| \leq |y|$ but not $\Leftarrow$ in general.

Let $X$ be a convex cone of $L^\infty(\mathcal{F})$. In what follows, we set $X = L^\infty(\mathcal{F})$ unless otherwise stated. An *error functional* on $X$ is a mapping $\Phi : X \times X \to \Re$ which satisfies the following six axioms.[2]

E1. $(\forall x, z \in X)\ \ \Phi(x, z) \geq 0$ ;

E2. $\Phi(x, z) = 0 \Leftrightarrow x = z$ ;

E3. $x - z \preceq x - z' \Rightarrow \Phi(x, z) \leq \Phi(x, z')$ ;

E4. $(\forall x \in X)\ \ \Phi(x, \cdot)$ is a convex function ;

E5. $(\forall x, z_1, z_2 \in X)(\forall \lambda \in (0,1))$
$z_1 \neq z_2$[3] and $(\exists \mathcal{G} \in \mathcal{F}^*)\ z_1, z_2 \in L^\infty(\mathcal{G})$ and $x \notin L^\infty(\mathcal{G})$
$\Rightarrow \Phi(x, \lambda z_1 + (1-\lambda) z_2) < \lambda \Phi(x, z_1) + (1-\lambda)\Phi(x, z_2)$ ; and

E6. $(\forall x \in X)(\exists p \in [1, +\infty))\ \ \| z_n - z_0 \|_p \to 0 \Rightarrow \underline{\lim}_{n\to\infty} \Phi(x, z_n) \geq \Phi(x, z_0)$ ,

where $\| \cdot \|_p$ is the $L^p$-norm. The number given by $\Phi(x, z)$ measures the error when we approximate $x$ by $z$. E1 and E2 are normalization axioms which are imposed so that the best approximation of $x$ is achieved by $x$ itself with no error. E3 requires consistency of the error with the "closeness" discussed earlier. E4-E6 are technical axioms which are sufficient for the existence of the best approximation. E4 and E5 require that $\Phi(x, \cdot)$ is convex and strictly convex, and E6 requires that $\Phi(x, \cdot)$ is strongly lower semi-continuous. While $\Phi$ resembles a metric, we do *not* impose the symmetric axiom: $\Phi(x, z) = \Phi(z, x)$.

We can construct an error functional from the functional $\widehat{\Phi} : X \to \Re$ which satisfies:

---

[2] The domain of $\Phi$ implicitly requires that $\Phi(x', z') = \Phi(x, z)$ if $x' = x$ a.e. and $z' = z$ a.e.

[3] $z_1 \neq z_2$ means that there does not exist $N$ such that $\{\omega \mid z_1(\omega) \neq z_2(\omega)\} \subset N$ and $\mu(N) = 0$.

N1. $(\forall x \in X)\ \ \widehat{\Phi}(x) \geq 0$ ;

N2. $\widehat{\Phi}(x) = 0 \Leftrightarrow x = 0$ ;

N3. $x \preceq y \Rightarrow \widehat{\Phi}(x) \leq \widehat{\Phi}(y)$ ;

N4. $(\forall x \in X)\ \ \lambda \geq 0 \Rightarrow \widehat{\Phi}(\lambda x) = \lambda \widehat{\Phi}(x)$ ;

N5. $(\forall x, y \in X)\ \ \widehat{\Phi}(x+y) \leq \widehat{\Phi}(x) + \widehat{\Phi}(y)$ ;

N6. $\widehat{\Phi}(x+y) < \widehat{\Phi}(x) + \widehat{\Phi}(y)$ unless $(\exists c > 0)\ x = cy$ ; and

N7. $(\exists p \in [1, +\infty))\ \ \| x_n - x_0 \|_p \to 0 \Rightarrow \underline{\lim}_{n\to\infty} \widehat{\Phi}(x_n) \geq \widehat{\Phi}(x_0)$ .

N7 can be replaced by the following axiom:

$$\text{N7}'. \quad (\exists p \in [1, +\infty) \text{ and } K > 0)(\forall x)\ \ \widehat{\Phi}(x) \leq K \|x\|_p .$$

It immediately follows that N5 and N7′ imply N7. By N4 and N5, $\widehat{\Phi}$ is a Minkowski functional. While $\widehat{\Phi}$ resembles a norm, it is *not* symmetric in the sense that $\widehat{\Phi}(-x) \neq \widehat{\Phi}(x)$.

**Lemma 2.1:** *Define* $\Phi(x, z) \equiv \widehat{\Phi}(x - z)$. *If* $\widehat{\Phi}$ *satisfies N1-N7, then* $\Phi$ *satisfies E1-E6.* (Proof in Appendix)

Given an error functional $\Phi$, we define

$$(\forall x \in X)(\forall \mathcal{G} \in \mathcal{F}^*)\quad M_\Phi(x|\mathcal{G}) = \arg\min\{\, \Phi(x, z) \,|\, z \in L^\infty(\mathcal{G}) \,\}. \tag{4}$$

The class of functions $M_\Phi(x|\mathcal{G})$, if it exists, is the "best" approximation to $x$ within the almost $\mathcal{G}$-measurable functions where approximation errors are measured by the error functional $\Phi$. When $x$ is (almost) $\mathcal{G}$-measurable, $z = x$ uniquely attains the minimum by E1 and E2. When $x$ is not (almost) $\mathcal{G}$-measurable, the next proposition guarantees the existence of the best approximation.

**Proposition 2.2:** *Let* $\Phi$ *be an error functional. Then for any* $x$ *and for any* $\mathcal{G}$, $M_\Phi(x|\mathcal{G})$ *is well-defined and unique (up to a* $\mu$*-equivalence class).* (Proof in Appendix)

We can always choose a version of $M_\Phi(x|\mathcal{G})$ which is $\mathcal{G}$-measurable, and refer to it as the *non-Hilbert conditioning*. The non-Hilbert conditioning (4) generalizes the conditional expectation (1), and coincides with it when $\Phi$ is specified by the $L^2$-metric. Any strictly increasing transformation of $\Phi$ can generate the same conditioning concept.

We define the *certainty equivalent* [4] generated by an error functional $\Phi$ by

$$(\forall x \in X)\quad M_\Phi(x) = M_\Phi(x|\{\phi, \Omega\}). \tag{5}$$

[4]The term "certainty equivalent" is used for $M$ only to mean that $M(x) = x$ when $x$ is constant. Recall that $x$ is now a random utility rather than a rondom income.

The $\mu$-equivalence class of functions $M_\Phi(x)$ is the "best" approximation to $x$ within the (almost) constant functions where approximation errors are measured by $\Phi$. Again, we can choose a version of $M_\Phi(x)$ which is constant. The certainty equivalent $M_\Phi(\cdot)$ generalizes the expectation functional $E(\cdot)$, and coincides with it when $\Phi$ is specified by the $L^2$-metric. Note that the equation (5) is the *definition* while the equation (2) is an *implication* of the conditional expectation. Given a certainty equivalent $M$, we can develop the concept of conditioning for $M$ by seeking for an error functional $\Phi$ such that $(\forall x)\ M(x) = M_\Phi(x|\{\phi, \Omega\})$. Such a $\Phi$ then would be used to define the conditioning of $M$ against any $\mathcal{G} \in \mathcal{F}^*$ by means of (4). Section 3 develops the conditioning for a broad class of non-expectation certainty equivalents in this way. Henceforth, we suppress $\Phi$ and simply write as $M(x|\mathcal{G})$ and $M(x)$ when $\Phi$ is understood.

An error functional $\Phi$ is of *integral form* if there exist mappings $\phi : \Re^2 \to \Re$ and $\varphi : \Re \to \Re$ such that

$$(\forall x, z \in X)\quad \Phi(x,z) = \varphi\left(\int_\Omega \phi(x(\omega), z(\omega))\mu(d\omega)\right)$$

and if $\Phi$ satisfies E1-E6. The error functionals of this form allow convenient characterization of $M(x|\mathcal{G})$ and $M(x)$, which follows immediately from the definition.

**Lemma 2.3:** *Let $\Phi$ be of integral form. If $\phi$ is continuously differentiable in its second argument and if $\varphi$ is strictly increasing, then, for any $x \in X$, $M(x)$ is defined as the unique solution $z$ to*

$$\int_\Omega \phi_2(x,z)\, d\mu = 0$$

*and $M(x|\mathcal{G})$ satisfies $(\forall h \in L^\infty(\mathcal{G}))$*

$$\int_\Omega \phi_2(x, M(x|\mathcal{G}))h\, d\mu = 0\,.$$

**Lemma 2.4:** *Let $\Phi$ be of integral form. If $\phi$ is twicely continuously differentiable and if $\varphi$ is strictly increasing, then $M(\cdot)$ is Gateaux differentiable and $(\forall x, h \in X)$*

$$\partial M(x;h) = -\frac{\int_\Omega \phi_{21}\,(x, M(x))\, h\, d\mu}{\int_\Omega \phi_{22}\,(x, M(x))\, d\mu}\,.$$

For each $x \in X$, the *value of information* $\mathcal{G}$ is a real number $V_x(\mathcal{G})$ defined by

$$V_x(\mathcal{G}) \equiv M(M(x|\mathcal{G})) - M(x)\,.$$

Here $M(M(x|\mathcal{G}))$ is the current "mean" utility when the agent expects partial information $\mathcal{G}$ to be obtained in the future, while $M(x)$ is the current "mean" utility when the agent expects no

more information. When $M$ coincides with the expectation $E$, $V_x(\mathcal{G}) = 0$ for any $x$ and for any $\mathcal{G}$ because of (3). However, $V$ could be either positive or negative in general because

$$M(x) \neq M(M(x|\mathcal{G}))\,. \tag{6}$$

The "biasedness" (6) of non-Hilbert conditioning contrasts with the "unbiasedness" (3) of conditional expectation. The agent with the $L^2$ error functional is indifferent to the partial information she might attain in the future (apart from the benefits of better planning which would become available with the additional information). On the other hand, with the non-Hilbert conditioning, the agent may strictly prefer to be uninformed. The preference over information can be introduced only by the "biasedness" of non-Hilbert conditioning. This new feature of non-Hilbert conditioning will be further illustrated by the examples in Sections 3 and 4.

## 3. Examples

This section considers several examples of non-expectation certainty equivalent, for each of which we develop the concept of conditioning. We also see its implications to the value of information.

### A. $L^p$ Error Functional

Let $p \in (1, +\infty)$. The *$L^p$ error functional* is defined by:

$$\Phi(x, z) \equiv \left( \int_\Omega \phi(x(\omega), z(\omega))\, d\mu(\omega) \right)^{1/p}, \quad \text{where} \quad \phi(x, z) = |x - z|^p\,.$$

This is the metric generated by $L^p$-norm and clearly satisfies E1-E6. We need to assume that $p \in (1, +\infty)$ since $M(x|\mathcal{G})$ is not unique in general when $p = 1$ or $+\infty$. When $p = 2$, the conditional expectation will be recovered. Dependent upon the value of $p$, the certainty equivalent generated by this error functional exhibits different sensitivity to events' probabilities. $M$ is very sensitive to probabilities when $p$ is close to one, while large $p$ shows the insensitivity of $M$ to probabilities. To see this, let $\Omega = \{\omega_1, \omega_2\}$, and let $(\forall i)\ p_i \equiv \mu(\{\omega_i\}) > 0$. When $p = 1$, $M(x) = x(\omega_i)$, where $i = \arg\max_j p_j$. (When there is a tie, $M(x)$ is any convex combination of such $x(\omega_i)$'s.) On the other hand, when $p = +\infty$, $M(x) = (1/2)(x(\omega_1) + x(\omega_2))$ regardless of $p_i$'s.

For $p \neq 2$, $M(x) \neq M(M(x|\mathcal{G}))$ in general as the following simple example shows. Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$; $(\forall i)\ p_i \equiv \mu(\{\omega_i\})$; $\mathcal{G} = \{\phi, \{\omega_1\}, \{\omega_2, \omega_3\}, \Omega\}$; $x_1 \equiv x(\omega_1) = x(\omega_2) < x_3 \equiv x(\omega_3)$; and $\gamma \equiv 1/(p-1)$. Then

$$M(x) = \frac{(p_1 + p_2)^\gamma}{(p_1 + p_2)^\gamma + p_3^\gamma} x_1 + \frac{p_3^\gamma}{(p_1 + p_2)^\gamma + p_3^\gamma} x_3 \quad \text{and}$$

$$M(M(x|\mathcal{G})) = \frac{p_1^\gamma(p_2^\gamma + p_3^\gamma) + p_2^\gamma(p_2+p_3)^\gamma}{(p_2^\gamma + p_3^\gamma)(p_1^\gamma + (p_2+p_3)^\gamma)} x_1 + \frac{p_3^\gamma(p_2+p_3)^\gamma}{(p_2^\gamma + p_3^\gamma)(p_1^\gamma + (p_2+p_3)^\gamma)} x_3 .$$

The several lines of algebra show that

$$M(x) \geq M(M(x|\mathcal{G})) \Leftrightarrow p \geq 2 .$$

Although this example is very specific and general cases seem to be very complicated, some intuition can be derived from this example. First, assume that $p > 2$. For this case, an intermediate aggregation (*i.e.*, $M(x|\mathcal{G})$) dampens (resp. promotes) the mean when the aggregation takes place among better (resp. worse) outcomes. This is because the values of outcomes (rather than the associated probabilities) count more and the best (resp. worst) outcome is always averaged downward (resp. upward). In the above example, the intermediate aggregation lowers the best outcome and hence dampens the mean. Next, assume that $p < 2$. For this case, an intermediate aggregation (*i.e.*, $M(x|\mathcal{G})$) promotes (resp. dampens) the mean when the relatively high probabilities are associated with the lower (resp. higher) outcomes. This is because the increase of the probability of better (resp. worse) events by the aggregation may cause the reversal of relative probabilities between better outcomes and worse outcomes. In the above example, the aggregation causes the increase in the probability of better outcome, which contributes increase of the mean.

### B. Asymmetric Error Functional

Let $p \in (1, +\infty)$ and let $\gamma > 0$. The *asymmetric error functional* is defined by:

$$\Phi(x,z) \equiv \left( \int_\Omega \phi(x(\omega), z(\omega))\, d\mu(\omega) \right)^{1/p}, \quad \text{where} \quad \phi(x,z) = \begin{cases} \gamma|x-z|^p & \text{if} \quad x \geq z \\ |x-z|^p & \text{if} \quad x < z . \end{cases}$$

When $\gamma < 1$, the certainty equivalent generated by this error functional is a parametric specification of preferences studied by Gul (1991), and exhibits *disappointment aversion.* This error functional measures the approximation errors asymmetrically. That is, an underestimate is tolerated (resp. penalized) compared with an overestimate when $\gamma < 1$ (resp. $\gamma > 1$). This error functional can be generated from the asymmetric "norm" which is defined in the next lemma.

**Lemma 3.1:** *Let $p \in (1, +\infty)$ and let $a, b > 0$. Define $\gamma : \Re \to \Re$ by*

$$(\forall x \in \Re) \quad \gamma(x) = \begin{cases} a & \text{if} \quad x \geq 0 \\ b & \text{if} \quad x < 0 . \end{cases}$$

*Finally, define $\|x\|_\gamma$ by*

$$\|x\|_\gamma = \left( \int_\Omega \gamma(x)|x|^p \, d\mu \right)^{1/p} .$$

*Then* $(\forall x, z \in X)$ $\|x+z\|_\gamma \leq \|x\|_\gamma + \|z\|_\gamma$. *Furthermore,* $\|x+z\|_\gamma < \|x\|_\gamma + \|z\|_\gamma$ *unless* $(\exists c > 0)$ $x = cz$ . (Proof in Appendix)

**Proposition 3.2:** *The asymmetric error functional is well-defined.*

**Proof:** By Lemma 2.1, it suffices to prove that $\widehat{\Phi}(x) = \|x\|_\gamma$ satisfies N1-N7. N1-N4 are obvious. N5 and N6 follow from Lemma 3.1. N7′ holds with $K = (\max\{a,b\})^{1/p}$. □

When $p = 2$, this error functional has a convenient implication on the value of information as the next proposition shows.

**Proposition 3.3:** *Let* $M$ *be generated by the asymmetric error functional with* $p = 2$. *Then for any* $\mathcal{G}$ *and for any* $x$, $M(x) \geq M\left(M(x|\mathcal{G})\right) \Leftrightarrow \gamma \leq 1$. (Proof in Appendix)

### C. Quasilinear Error Functional

For this example, set $X$ to be the positive cone of $L^\infty(\mathcal{F})$, and let $\alpha > 0$. The *quasilinear error functional* is defined by:

$$\Phi(x,z) \equiv \int_\Omega \phi(x(\omega), z(\omega))\, d\mu(\omega), \quad \text{where} \quad \phi(x,z) = \alpha x^{\alpha+1} - (\alpha+1)x^\alpha z + z^{\alpha+1}.$$

When $\alpha = 1$, the conditional expectation will be recovered. The certainty equivalent generated by this error functional is a parametric family of quasilinear mean, $(E[x^\alpha])^{1/\alpha}$.

**Proposition 3.4:** *The quasilinear error functional is well-defined.*

**Proof:** This is a special case of Proposition 3.5. □

By Lemma 2.3, we can compute $M(x|\mathcal{G})$ explicitly as $(\forall x)(\forall \mathcal{G})\ M(x|\mathcal{G}) = (E[x^\alpha|\mathcal{G}])^{1/\alpha}$. Hence, for this error functional, the value of information is always 0 because $M(M(x|\mathcal{G})) = (E[E[x^\alpha|\mathcal{G}]])^{1/\alpha} = (E[x^\alpha])^{1/\alpha} = M(x)$.

### D. Implicit Error Functional

For this example, we first define the certainty equivalent, and then seek for the error functional which generates this certainty equivalent. Given $x \in X$, $M(x)$ is defined as the unique solution $z \in \Re$ to

$$\int_\Omega \varphi(x(\omega), z)\, d\mu(\omega) = 0$$

where $\varphi : \Re^2 \to \Re$ satisfies: (i) $\varphi(x,x) = 0$, (ii) $(\forall x)$ $\varphi(x,\cdot)$ is strictly decreasing, and (iii) $(\forall x)$ $\varphi(x,\cdot)$ is continuous. Note that $z = M(x)$ is well-defined by the mean value theorem. This certainty equivalent, called implicit mean, was studied by Fishburn (1986). By the continuity

(iii), there exists $\widehat{\phi} : \Re^2 \to \Re$ such that $(\forall x, z)\ \widehat{\phi}_2(x, z) = -\varphi(x, z)$. Furthermore, $\widehat{\phi}(x, \cdot)$ is strictly convex by (ii) and has a minimum at $z = x$ by (i). Finally, let $\phi(x, z) \equiv \widehat{\phi}(x, z) - \widehat{\phi}(x, x)$. We define the *implicit error functional* by:

$$\Phi(x, z) \equiv \int_\Omega \phi(x(\omega), z(\omega))\, d\mu(\omega)\,.$$

Note that the quasilinear error functional is recovered when $\varphi(x, z) = (\alpha + 1)x^\alpha - (\alpha + 1)z^\alpha$.

**Proposition 3.5:** *The implicit error functional is well-defined, and it generates the implicit mean $M(x)$.*

**Proof:** E1-E6 follow since $\phi(x, \cdot)$ is a strictly convex function which has the minimized value 0 at $z = x$. E7 holds by Aubin and Ekeland (1984, p. 13) since $\phi(x, \cdot)$ is continuous. The second statement in the proposition is obvious by Lemma 2.3. □

By the construction of $\Phi$ and Proposition 3.5, we can always find the error functional which generates the given implicit mean. The value of information is non-zero except when $\varphi$ is separable (as in Example C).

### E. Rank-dependent Error Functional

Let $\theta$ be a normalized capacity on $(\Omega, \mathcal{F})$, that is, $\theta : \mathcal{F} \to [0, 1]$ is a mapping which satisfies $\theta(\phi) = 0$, $\theta(\Omega) = 1$, and $A \subset B \Rightarrow \theta(A) \leq \theta(B)$. We will also assume that $\theta$ is *convex* in the sense that $\theta(A \cup B) + \theta(A \cap B) \geq \theta(A) + \theta(B)$. A simple example of a convex capacity is $(\forall A)\ \theta(A) = (\mu(A))^\alpha$, where $\alpha \geq 1$. Let $X$ be the positive cone of $L^\infty(\mathcal{F})$. A *Choquet integral* with respect to a capacity $\theta$ is defined by:

$$(\forall x \in X) \quad \int_\Omega x(\omega)\, \theta(d\omega) \equiv \int_0^\infty \theta(\,\{\omega \in \Omega | x(\omega) \geq t\}\,) dt\,,$$

where the integral in the right-hand side is an improper Riemann (or equivalently, Lebesgue) integral. The Choquet integral crucially depends upon the ranking of outcomes, and (given $\theta$'s convexity) it captures the notion of *uncertainty aversion* (for example, see Gilboa, 1987, Schmeidler, 1989, and Chateauxneuf, 1991).

We construct the error functional which generates the certainty equivalent defined by the Choquet integral. To this end, we introduce a couple of concepts. We denote by *ba* the space of bounded charges. That is, *ba* is the space of bounded finitely additive set functions on $(\Omega, \mathcal{F})$ which are null at $\phi$ and absolutely continuous with respect to $\mu$. Note that $ba = (L^\infty(\mathcal{F}))^*$, the topological dual of $L^\infty(\mathcal{F})$. Let $\langle \varphi_n \rangle$ be a nondecreasing sequence of simple functions which uniformly converges to $x$. Such a sequence always exists. Then the *Dunford-Schwartz integral*

of $x \in X$ with respect to $\nu \in ba$ is defined by

$$\int_\Omega x(\omega)\,\nu(d\omega) \equiv \lim_{n\to\infty} \int_\Omega \varphi_n(\omega)\,\nu(d\omega)\,,$$

where the integrals of simple functions in the RHS are defined in the same manner as for an integral with respect to a measure. The limit exists and is independent of the choice of the sequence of simple functions (Danford and Schwartz, 1954, p. 111). Finally, the *core* of the normalized capacity $\theta$ is defined by

$$\text{core}(\theta) = \{\, \nu \in ba \mid (\forall A \in \mathcal{F})\ \theta(A) \le \nu(A) \ \text{ and } \ \nu(\Omega) = 1 \,\}\,.$$

It immediately follows that core($\theta$) is weak * compact. The following lemma is due to Schmeidler (1986):

**Lemma 3.6:** *Let $\theta$ be a normalized capacity. Then $\theta$ is convex if and only if*

$$(\forall x \in X) \quad \int_\Omega x\,d\theta = \min\left\{ \int_\Omega x\,d\nu \;\middle|\; \nu \in \text{core}(\theta) \right\}.$$

We now define the error functional. Let $\theta$ be a normalized capacity which is convex. Then define a mapping $x \mapsto \nu_x$ from $X$ into core($\theta$) such that

$$\int_\Omega x\,d\theta = \int_\Omega x\,d\nu_x\,,$$

where the RHS is the Dunford-Schwartz integral. Such a mapping exists by Lemma 3.6 while it is not unique in general. We define the *rank-dependent error functional* by:

$$\Phi(x,z) \equiv \int_\Omega (x(\omega) - z(\omega))^2\,d\nu_x(\omega)\,.$$

Note that the integral is again in the sense of Dunford and Schwartz since $\nu_x$ is a charge rather than a measure. We can prove

**Proposition 3.7:** *The rank-dependent error functional is well-defined, and it generates the Choquet integral.* (Proof in Appendix)

For this error functional, the value of information is in general non-zero. Lemma 2.3 (and the similar argument to the proof of Proposition 3.7) shows

$$M(x) = \int_\Omega x\,d\nu_x = \int_\Omega M(x|\mathcal{G})\,d\nu_x\,,$$

but the last expression does not equal $M(M(x|\mathcal{G}))$ since $\nu_x$ and $\nu_{M(x|\mathcal{G})}$ are in general different.

## 4. Application to 2-Period Dynamic Model: An Example

The general two-period model is described as follows. Let $(\Omega, \mathcal{F}, \mu)$ be the set of all possible states of the world of tomorrow together with all conceivable events and a probability measure on it. And let $c_0$ and $c_1$ be a consumption of today and of tomorrow, respectively. $c_0$ is a non-negative real and $c_1 : \Omega \to \Re_+$ is a function of tomorrow's state. The feasibility of a consumption plan is described by the compact-valued and $\mathcal{F}$-measurable correspondence $F : \Omega \twoheadrightarrow \Re_+^2$. It will be required that $(\forall\omega)\ (c_0, c_1(\omega)) \in F(\omega)$. A function $u : \Omega \to \Re$ is a $\mathcal{F}$-measurable utility function which gives future's utility given tomorrow's state $\omega$. A function $W : \Re_+ \times \Re \to \Re$ is an intertemporal aggregator which generates today's utility given today's consumption and tomorrow's "mean" utility. In order to caluculate tomorrow's "mean" utility, the agent employs the certainty equivalent $M$ which is generated by some error functional.

The information structure of the model is as follows. Let $\mathcal{F}^*$ be the space of sub-$\sigma$-algebras of $\mathcal{F}$, and let $\mathcal{G}^*$ be some subset of $\mathcal{F}^*$. Today, the agent chooses any $\mathcal{G} \in \mathcal{G}^*$ together with a feasible consumption plan $(c_0, c_1)$. The information $\mathcal{G}$ will be revealed to the agent tomorrow, and $c_1$ must be chosen so as to be $\mathcal{G}$-measurable. Therefore, the finer $\mathcal{G}$ is, the more variety of functions the agent can consider as a possible plan of tomorrow's consumption $c_1$. This substantiates the idea that the additional information contributes the utility by allowing the more flexible planning. Finally, the utility of tomorrow will be computed as a conditional utility given $\mathcal{G}$, and this conditional utility will be aggregated again over states by $M$ to generate "mean" utility of tomorrow. When $M$ is given by the expectation operator $E$ and when $W$ is linear in its second argument, this double aggregation leads to the identical "mean" utility regardless of $\mathcal{G}$. But since $M$ need not be $E$ now, the "mean" utility may be affected by the choice of $\mathcal{G}$.

In a summary, the agent chooses $c_0$, $c_1$, and $\mathcal{G}$ to maximize:

$$W\left(c_0, M\left[W(c_1(\omega), M[u(\omega)|\mathcal{G}])\right]\right) \tag{7}$$

$$\text{subject to} \begin{cases} (\forall\omega \in \Omega)\ \ (c_0, c_1(\omega)) \in F(\omega) \\ \mathcal{G} \in \mathcal{G}^* \ \text{ and} \\ c_1 \ \text{ is } \mathcal{G}\text{-measurable.} \end{cases}$$

When $W(c, m) = v(c) + \beta m$ and $M = E$, where $v : \Re_+ \to \Re$, $\beta > 0$ and $E$ is an expectation operator, equation (7) is reduced to

$$\begin{aligned} & v(c_0) + \beta E[v(c_1(\omega))] + \beta^2 E[\,E[u(\omega)|\mathcal{G}]\,] \\ = \ & v(c_0) + \beta E[v(c_1(\omega))] + \beta^2 E[u(\omega)] \end{aligned}$$

Here $\mathcal{G}$ vanishes, and hence it cannot affect the overall utility. More information is always preferred. Next suppose that $W(c,m) = m$. Eqaution (7) is now reduced to

$$M[\,M[u(\omega)|\mathcal{G}]\,]$$

Hence if $M \neq E$, $\mathcal{G}$ could affect the overall utility even for this simple intertemporal aggregator.

A finer $\mathcal{G}$ leads to better planning, but it may not be preferred according to the *anti-*preference for information itself. To illustrate this point, we now provide a specific example of the above model in which the attainment of new information (and hence more flexible planning) may not necessarily improve the agent's overall utility.

*Information Structure.* Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$; $\mathcal{F} = 2^{\Omega}$; $(\forall i)\ p_i \equiv \mu(\{\omega_i\})$; $\mathcal{G}_0 \equiv \{\phi, \Omega\}$; $\mathcal{G}_1 \equiv \{\phi, \{\omega_1\}, \{\omega_2, \omega_3\}, \Omega\}$; and $\mathcal{G}^* = \{\mathcal{G}_0, \mathcal{G}_1\}$. For this simple example, the agent has only two alternatives: no information at all or partial information which can't distinguish state 2 from state 3. We assume that it takes no cost obtaining information.

*Feasibility.* Let $\varepsilon > 0$ and let $\bar{c} > 0$. The feasibility correspondence is given by:

$$(\forall\omega)\quad F(\omega) = \begin{cases} \{(x_0, x_1) \in \Re^2_+ \,|\, x_0 = 0 \text{ and } 0 \leq x_1 \leq \bar{c}\} & \text{if} \quad \omega = \omega_1 \\ \{(x_0, x_1) \in \Re^2_+ \,|\, x_0 = 0 \text{ and } 0 \leq x_1 \leq \bar{c} + \varepsilon\} & \text{if} \quad \omega = \omega_2 \text{ or } \omega_3 \end{cases}$$

When $\mathcal{G}_0$ is chosen by the agent, the best consumption available tomorrow is given by $(\forall\omega)\ c_1(\omega) = \bar{c}$. When $\mathcal{G}_1$ is chosen by the agent, the best consumption available tomorrow is given by $c_1(\omega_1) = \bar{c}$ and $c_1(\omega_2) = c_1(\omega_3) = \bar{c} + \varepsilon$. The more information leads to a better consumption of tomorrow as we expect.

*Preference.* Let $b > 0$. Define $W$ by $(\forall c, m)\ W(c,m) = c + m$ and $u$ by

$$(\forall\omega)\quad u(\omega) = \begin{cases} 0 & \text{if} \quad \omega = \omega_1 \text{ or } \omega_2 \\ b & \text{if} \quad \omega = \omega_3 \end{cases}$$

The state 3 is special in the sense that the agent gets a bonus utility $b$ only when this state is realized. Finally, the certainty equivalent $M$ is the one generated by the $L^p$ error functional (which was introduced in Section 3.A). Under these specifications, equation (7) is reduced to $M(\,c_1(\omega) + M(u(\omega)|\mathcal{G})\,)$.

Let $\gamma = 1/(p-1)$. The overall maximum utility when the agent chooses $\mathcal{G}_0$ is given by:

$$U_0 \;\equiv\; \bar{c} + M(u(\omega)) \;=\; \bar{c} + \frac{p_3^{\gamma}}{(p_1+p_2)^{\gamma} + p_3^{\gamma}} b\,.$$

On the other hand, the overall maximum utility when the agent chooses $\mathcal{G}_1$ is given by:

$$U_1 \;\equiv\; M(\,c_1(\omega) + M(u(\omega)|\mathcal{G})\,) \text{ where } c_1(\omega) = \begin{cases} \bar{c} & \text{if} \quad \omega = \omega_1 \\ \bar{c} + \varepsilon & \text{if} \quad \omega = \omega_2 \text{ or } \omega_3 \end{cases}$$

$$= \bar{c} + \frac{(p_2+p_3)^{\gamma}}{(p_1^{\gamma} + (p_2+p_3)^{\gamma})}\varepsilon + \frac{p_3^{\gamma}(p_2+p_3)^{\gamma}}{(p_2^{\gamma}+p_3^{\gamma})(p_1^{\gamma}+(p_2+p_3)^{\gamma})} b\,.$$

As the example in Section 3.A shows, $U_0 = M(x) \leq M(M(x'|\mathcal{G})) = U_1$ when $p \leq 2$. In this case, obtaining partial information merits the higher utility (partly because of the better planning and partly because of the preference for information itself). On the contrary, when $p > 2$ and $\varepsilon$ is small enough (or, $b$ is large enough), $U_0 = M(x) > M(M(x'|\mathcal{G})) = U_1$. For instance, when $p > 2$ (and hence $\gamma < 1$) and $p_1 = p_2 = p_3$, it immediately follows that $U_0 > U_1$ if (and only if) $b > [2^\gamma/(1 - 2^{\gamma-1})]\varepsilon$. In this case, the agent prefers to remain uninformed rather than become informed. This is because the *anti*-preference for information dominates the merit by the better planning.

## 5. Concluding Remarks

The simple example of 2-period model in Section 4 shows that the agent with a general error functional (which may be different from $L^2$-norm) can exhibit a strict preference or a strict *anti*-preference for information, the latter of which sometimes dominates even the merit of better planning. In this manner, the non-Hilbert conditioning allows us to make more general models in which the information will be jurisdictionally chosen by the agent and hence will be endogenized. In the traditional dynamic models, the accumulation of agent's information is exogenously given by some filtration of $\sigma$-algebras. In another word, the choice of variables the agent observes to obtain information is out of her control (where the filtration is generated by the variables). Multi-period (in particular, infinite-horizon) models with the non-Hilbert conditioning contrast with the traditional ones and would provide new economic implications by endogenizing the problem of which variables to observe. In an attempt to analyze such models (including the general 2-period model described at the beginning of Section 4), it would be necessary to study the existence of the optimal information level, which in turn requires the "continuity" of $M(x|\cdot)$ with respect to some information topology on $\mathcal{F}^*$. To this end, we would need to extend the argument developed by Allen (1983), which employs Boylan's (1971) topology for information.

# APPENDIX

**Proof of Lemma 2.1:** The following implications are immediate: N1 $\Rightarrow$ E1; N2 $\Rightarrow$ E2; N3 $\Rightarrow$ E3; and N7 $\Rightarrow$ E6. N4 and N5 imply E4 since:

$$\begin{aligned}(\forall \lambda \in (0,1)) \qquad & \Phi(x, \lambda z_1 + (1-\lambda) z_2) \\ &= \widehat{\Phi}(x - (\lambda z_1 + (1-\lambda) z_2)) \\ &= \widehat{\Phi}(\lambda(x - z_1) + (1-\lambda)(x - z_2)) \\ &\leq \widehat{\Phi}(\lambda(x - z_1)) + \widehat{\Phi}((1-\lambda)(x - z_2)) \\ &= \lambda \widehat{\Phi}(x - z_1) + (1-\lambda)\widehat{\Phi}(x - z_2) \\ &= \lambda \Phi(x, z_1) + (1-\lambda)\Phi(x, z_2)\,.\end{aligned} \tag{8}$$

Finally we prove E5. The inequality in (8) is strict unless there exists $c \in \Re_+$ such that $\lambda(x - z_1) = c(1-\lambda)(x - z_2)$, which is the case only when $\lambda - c(1-\lambda) = 0$ or $x$ is a linear combination of $z_1$ and $z_2$. But the both are impossible under the presupposition of E5. □

**Proof of Proposition 2.2:** Let $x \in X$. Then there exists $b \in \Re_+$ such that $|x| \leq b$. Define

$$(L^\infty(\mathcal{G}))_b \equiv \{\, x \in L^\infty(\mathcal{G}) \,|\, \operatorname{ess\,sup}_{\omega \in \Omega} |x(\omega)| \leq b \,\}\,.$$

We can look for $z$ in $(L^\infty(\mathcal{G}))_b$ because $x - (z \wedge b) \vee (-b) \preceq x - z$ and because it implies $\Phi(x, (z \wedge b) \vee (-b)) \leq \Phi(x, z)$ by E3.

Let $\langle z_n \rangle_{n=1}^\infty$ be a sequence in $(L^\infty(\mathcal{G}))_b$ such that

$$\Phi(x, z_n) \to \inf_z \Phi(x, z) \equiv \inf\{\, \Phi(x, z) \,|\, z \in (L^\infty(\mathcal{G}))_b \,\}\,. \tag{9}$$

Since $(L^\infty(\mathcal{F}))_b$ is weak * compact, there exists a subsequence $\langle z_{n_i} \rangle_{i=1}^\infty$ and $z_0 \in (L^\infty(\mathcal{F}))_b$ such that $z_{n_i}$ converges to $z_0$ in the weak * topology. We will show that $z_0 \in (L^\infty(\mathcal{G}))_b$ in the rest of this paragraph. Fix $p$ with which $\Phi$ meets E6 and let $q$ be its conjugate (When $p = 1$, let $q = +\infty$). We claim that $z_{n_i}$ converges to $z_0$ in the weak topology of $L^p(\Omega, \mathcal{F}, \mu)$. To see this, note that $z_{n_i}$ and $z_0$ live in $L^p$, and that, by the definition of the weak * topology,

$$(\forall x \in L^1) \quad \int_\Omega x z_{n_i}\, d\mu \to \int_\Omega x z_0\, d\mu\,.$$

The claim follows because $L^q \subset L^1$ and $L^q = (L^p)^*$. Therefore, some sequence of convex combinations of the elements $z_{n_i}$ converges to $z_0$ in the $L^p$-norm by the Mazur's lemma. Then some subsequence of this sequence converges to $z_0$ almost everywhere. Since each component of this subsequence is almost $\mathcal{G}$-measurable, $z_0$ is also almost $\mathcal{G}$-measurable.

This paragraph proves that $\Phi(x, z_0) = \inf_z \Phi(x, z)$. Because $z_{n_i}$ converges to $z_0$ in the weak topology by the second paragraph, and because $\Phi(x, \cdot)$ is lower semi-continuous in the weak topology by E4 and E6, $\underline{\lim}_{n \to \infty} \Phi(x, z_{n_i}) \geq \Phi(x, z_0)$, which implies $\Phi(x, z_0) = \inf_z \Phi(x, z)$ by (9).

Finally, we prove the uniqueness. If $x$ is almost $\mathcal{G}$-measurable, $z = x$ attains the unique minimum by E1 and E2. Assume that $x$ is not almost $\mathcal{G}$-measurable, and that almost $\mathcal{G}$-measurable distinct functions $z_1$ and $z_2$ both attain the minimum. But this contradicts E5. □

**Proof of Lemma 3.1:** If $x = 0$ or $z = 0$, the conclusion is trivial. Hence assume that $\|x\|_\gamma \equiv \alpha > 0$ and $\|z\|_\gamma \equiv \beta > 0$. Define $x_0 = x/\alpha$ and $z_0 = z/\beta$. Then $\alpha x_0 = x$ and $\|x_0\|_\gamma = 1$ (since $\|\cdot\|_\gamma$ is positively homogeneous). The same is true for $z_0$. If we define $\lambda = \alpha/(\alpha+\beta)$, then

$$\begin{aligned}
& \gamma(x+z)|x+z|^p \\
= \ & \gamma(\alpha x_0 + \beta z_0)|\alpha x_0 + \beta z_0|^p \\
= \ & \gamma(\lambda(\alpha+\beta)x_0 + (1-\lambda)(\alpha+\beta)z_0)|\lambda(\alpha+\beta)x_0 + (1-\lambda)(\alpha+\beta)z_0|^p \\
= \ & (\alpha+\beta)^p \gamma(\lambda x_0 + (1-\lambda)z_0)|\lambda x_0 + (1-\lambda)z_0|^p \\
\leq \ & (\alpha+\beta)^p \left(\lambda\gamma(x_0)|x_0|^p + (1-\lambda)\gamma(z_0)|z_0|^p\right),
\end{aligned}$$

where the last equality holds since $\alpha + \beta > 0$, and the inequality holds (with an equality only when $x_0 = z_0$) since $\gamma(\cdot)|\cdot|^p$ is strictly convex. Therefore,

$$\begin{aligned}
& \|x+z\|_\gamma^p \\
= \ & \int_\Omega \gamma(x+z)|x+z|^p \, d\mu \\
\leq \ & \int_\Omega (\alpha+\beta)^p \left(\lambda\gamma(x_0)|x_0|^p + (1-\lambda)\gamma(z_0)|z_0|^p\right) d\mu \\
= \ & (\alpha+\beta)^p \left(\lambda\|x_0\|_\gamma^p + (1-\lambda)\|z_0\|_\gamma^p\right) \\
= \ & (\alpha+\beta)^p .
\end{aligned}$$

By taking $p$−root, the proof is completed. □

**Proof of Proposition 3.3:** Define $\varphi : \Re^2 \to \Re$ by

$$\varphi(x,z) = \begin{cases} \gamma(x-z) & \text{if} \quad x \geq z \\ x - z & \text{if} \quad x < z \end{cases}$$

Then $M(x)$ is defined as a unique solution $z$ to $\int_\Omega \varphi(x(\omega), z)\, d\mu(\omega)$.

Let $x, h \in L^\infty(\mathcal{F})$ and let $\gamma > 1$. Then by the convexity of $\varphi$,

$$\begin{aligned}
(\forall\omega) \qquad & \varphi(x(\omega)+h(\omega), M(x+h)) - \varphi(x(\omega), M(x)) \\
& - \varphi_1^+(x(\omega), M(x))h(\omega) - \varphi_2^-(x(\omega), M(x))[M(x+h) - M(x)] \\
\geq \ & 0
\end{aligned}$$

where $\varphi_1^+$ and $\varphi_1^-$ are right-hand-side and left-hand-side derivatives, respectively. Hence

$$\int_\Omega \left(\varphi_1^+(x(\omega), M(x))h(\omega) + \varphi_2^-(x(\omega), M(x))[M(x+h) - M(x)]\right) d\mu \leq 0 .$$

By setting $h = M(x|\mathcal{G}) - x$,

$$M(M(x|\mathcal{G})) - M(x) \geq -\frac{\int_\Omega \varphi_1^+(x(\omega), M(x))[M(x|\mathcal{G}) - x]\, d\mu}{\int_\Omega \varphi_2^-(x(\omega), M(x))\, d\mu}.$$

Note that $\int_\Omega \varphi_2^-(x(\omega), M(x))\, d\mu \leq 0$, and that $\int_\Omega \varphi_1^+(x(\omega), M(x))[M(x|\mathcal{G}) - x]\, d\mu \geq 0$ because

$$\begin{aligned}
& \int_\Omega \varphi_1^+(x(\omega), M(x))[M(x|\mathcal{G}) - x]\, d\mu \\
= & \int_{\{x \geq M(x)\}} \gamma(M(x|\mathcal{G}) - x)\, d\mu + \int_{\{x < M(x)\}} (M(x|\mathcal{G}) - x)\, d\mu \\
= & (\gamma - 1) \int_{\{x \geq M(x)\}} (M(x|\mathcal{G}) - x)\, d\mu + \int_\Omega (M(x|\mathcal{G}) - x)\, d\mu \\
\geq & (\gamma - 1) \int_{\{x \geq M(x|\mathcal{G})\}} (M(x|\mathcal{G}) - x)\, d\mu + \int_\Omega (M(x|\mathcal{G}) - x)\, d\mu \\
= & 0,
\end{aligned}$$

where the last equality holds by Lemma 2.3. This completes the proof for $\gamma > 1$. The similar argument applies for $\gamma < 1$. □

**Proof of Proposition 3.7:** E1-E6 are obvious. Yoshida-Hewitt theorem claims $(\forall x)\ \nu_x = \nu_x^c + \nu_x^p$, where $\nu_x^c$ is a countably additive part of $\nu_x$ and $\nu_x^p$ is a pure charge, both of which are nonnegative. Hence, $\Phi(x, z) \geq \int_\Omega (x - z)^2 d\nu_x^c$, from which the lower semi-continuity of $\Phi$ in $z$ follows in a usual manner since $\nu_x^c$ is now a measure.

To show that $\Phi$ generates the Choquet integral, let $\langle z_n \rangle_n$ be a sequence of real numbers which converges to $z$. Then $((x(\omega) - z_n)^2 - (x(\omega) - z)^2)/(z_n - z)$ converges to $(d/dz)(w(\omega) - z)^2$ in "charge" because

$$(\forall \omega) \quad \left| \frac{(x(\omega) - z_n)^2 - (x(\omega) - z)^2}{z_n - z} - \frac{d}{dz}(w(\omega) - z)^2 \right| = |z_n - z|.$$

Therefore, by the dominated convergence theorem for charges (Rao and Rao, 1983, p. 131), $(d/dz)\int_\Omega (x(\omega) - z)^2\, d\nu_x = \int_\Omega (d/dz)(x(\omega) - z)^2\, d\nu_x$, which proves the claim by the definition of $\nu_x$. □

## References

Allen, Beth (1983): "Neighboring Information and Distributions of Agent's Characteristics under Uncertainty", *Journal of Mathematical Economics* 12, 63-101.

Aubin, Jean-Pierre and Ivar Ekeland (1984): *Applied Nonlinear Analysis*, Wiley-Interscience.

Billingsley, Patrick (1986): *Probability and Measure* (Second Edition), John Wiley and Sons, New York.

Boylan, Edward S. (1971): "Equiconvergence of Martingales", *The Annals of Mathematical Statistics* Vol. 42, No. 2, 552-559.

Chateauneuf, Alain (1991): "On the Use of Capacities in Modeling Uncertainty Aversion and Risk Aversion", *Journal of Mathematical Economics* 20, 343-369.

Dunford, Nelson and Jacob T. Schwartz (1954): *Linear Operators, Part I: General Theory*, Wiley-Interscience, London.

Fishburn, Peter C. (1986): "Implicit Mean Value and Certainty Equivalence", *Econometrica* 54, 1197-1205.

Gilboa, Itzhak (1987): "Expected Utility with Purely Subjective Non-Additive Probabilities", *Journal of Mathematical Economics* 16, 65-88.

Gul, Faruk (1991): "A Theory of Disappointment Aversion", *Econometrica* 59, 667-686.

Rao, K. P. S. Bhaskara and M. Bhaskara Rao (1983): *Theory of Charges*, Academic Press.

Schmeidler, David (1986): "Integral Representation without Additivity", *Proceedings of the American Mathematical Society* Vol. 97, No. 2, 255-261.

Schmeidler, David (1989): "Subjective Probability and Expected Utility without Additivity", *Econometrica* 57, 571-587