Practical Application and Discernability of Risky Choice Models

by

David R. Just
Cornell University

Richard E. Just
University of Maryland

Working Paper
(Not to be quoted without permission)

April 2009

Practical Application and Discernability of Risky Choice Models

Dozens of models of decision making under uncertainty have been proposed over the last several decades (see Starmer 2000 for a review). Some have even been resurrected from earlier eras (e.g., the safety rules of Roy 1952, Katoka 1963, and Telser 1955-6). All of these models have a common purpose of improving upon the standard von Neumann-Morgenstern expected utility model, specifically, in representing anomalies that appear to violate the predictions of expected utility theory.

Individuals can appear to violate von Neumann-Morgenstern expected utility maximization for two reasons: failure of preferences (dependence of utility of outcomes on the particular gambles being offered) or failure of perceptions (judgment bias). Preference failure has been addressed by various outcome weighting models that depart from a standard utility function while perception failure has been addressed with various probability weighting models. Models that modify evaluation of outcomes include Machina (1982), Yaari (1987), and Chew (1983, 1989). Models that modify probability weightings include Edwards (1955) and Quiggin (1982). Kahneman and Tversky's(1979) prospect theory, given a particular reference point, fall into the latter category.

Interestingly, models proposed to address these two types of failure appear to leave a nearly empty intersection. For example, Machina (1982), in his development of generalized expected utility theory, rejects the notion that individuals distort probabilities (mainly because many models of perception errors violate monotonicity). In contrast, Hey and Strazzera (1989) find evidence that individuals violate monotonicity when stochastic dominance is hard to identify. On the other hand, the judgment bias literature has documented a variety of types of perception failure including overconfidence (Alpert and Raiffa 1982; Camerer 1995; Lichtenstein, Frischhoff, and Phillips 1982; Murphy and Winkler 1974), availability bias (Tversky and Kahneman 1982), law of small numbers bias (Tversky and Kahneman 1982), representative bias (Tversky and Kahneman 1982), and conservatism (Edwards 1982). A common thread in this literature is that individuals give inadequate weight to rare events and difficult-to-process information, and excessive weight to common or recent events and easy-to-process information.

Given the substantial evidence demonstrating judgment bias (perception failure), the evidence that supports models with modified utility functions under the assumption that probabilities are not distorted is exceedingly weak. We demonstrate this weakness by showing that models of preference failure are mathematically equivalent to models of perception failure. For practical purposes, these models can be discerned using behavioral data alone only by imposing arbitrary assumptions. These results follow from the inability to identify econometrically, except by arbitrary parametric specifications, two multiplicative functions that have common variables (see Just and Just, forthcoming). We show that this implies the popular Arrow-Pratt measures of risk aversion are not truly identifiable for empirical purposes in common situations of data availability. In this context, we argue that the addition to knowledge of research on risk behavior is limited because models with vastly different risk behavior are empirically equivalent.

After exploring the extent of econometric discernment possible in typical empirical applications (defined as the case where behavioral data alone are available, whether from the real-world or laboratory settings), we suggest a practical empirical approach that focuses on truly identifiable concepts combining preferences and perceptions. This avoids the need for unverifiable assumptions to identify preferences and perceptions separately. The resulting empirical framework relaxes typical restrictions on the mean-variance model while focusing on the well-understood concepts of risk premium and certainty equivalent. This provides sufficient structure for policy and welfare analysis. Further, it provides analogs of Arrow-Pratt measures of risk aversion that are intuitively plausible and econometrically identifiable, but which have an accurate empirical interpretation independent on unverifiable underlying assumptions in the spirit of the recent literature on sufficient statistics in economics (Chetty 2008).

Both Preferences and Perceptions Affect Behavior under Risk

Most of the major theoretical models that have been proposed for decision making under risk have (or can be represented by) a common structure whereby the distribution of random outcomes is valued by summing or integrating some type of utility or valuation function with respect to some weighting of the probability distribution. The most popular of these models modify the classical von Neumann-Morgenstern expected utility model by modifying one of its two essential components. Some modify the utility function, which is ordinarily a function representing how outcomes of a risky choice are valued under certainty, by making the function dependent on the distribution. Others modify the weights implied by the probability distribution.

In some models the probability distribution is transformed to represent misperceptions, such as in Edwards' (1955) model of decision weights where distributions are valued by

$$v(F) = \int_{-\infty}^{\infty} x\pi\left(f(x)\right)dx$$

where $f(x)$ is the probability density of possible payoffs represented by x, and $\pi(\cdot)$ is a nonnegative weighting function. This model was later generalized (Edwards 1961) to reflect diminishing marginal utility using a utility function u,

$$v(F) = \int_{-\infty}^{\infty} u(x)\pi\left(f(x)\right)dx$$

where $u(\cdot)$ is a utility function with positive monotonicity. In other models, the probability distribution is modified by heuristics about how individuals process information about probabilities. For example, to the extent that such rules generate unique outcomes, the $\pi(\cdot)$ function in this model can represent the heuristic rules of prospect theory such as those that ignore minor differences in probabilities and common components of distributions.

In the rank-dependent utility model of Quiggin (1982), valuation of distributions depends on the set of possible outcomes such that

$$v(p) = \sum_{i=1}^{n} x_i \pi_i(p) \tag{1}$$

where $x_1,\ldots,x_n$ represents outcomes ordered from best to worst, $p_1,\ldots,p_n$ are their respective probabilities, and the rank-dependent probability function is defined by

$$\pi_i(p) = g(p_1 + \ldots + p_i) - g(p_1 + \ldots + p_{i-1}) \tag{2}$$

for some positive monotonic function $g(\cdot)$.

Other approaches focus on transformations of the utility function based on the probability or relative probability attached to the outcome, as in Chew's (1983) weighted utility model,

$$v(F) = \int_{-\infty}^{\infty} a(x)u(x)dF(x) \Big/ \int_{-\infty}^{\infty} a(x)dF(x), \tag{3}$$

where F is the distribution of possible payoffs and a is a non-vanishing utility weighting function. The generalized expected utility model of Machina considers a more general modification of the utility function dependent on the entire distribution whereby

$$v(F) = \int_{-\infty}^{\infty} u(x \mid F) dF(x) \tag{4}$$

with the motivations that the distribution of outcomes affects the valuation of specific outcomes and preferences over gambles seem to be independent of wealth. Additionally, some theories modify the utility function by making the value of the outcome dependent upon a local probability measure or reference point attached to the outcome. For example, the generalized prospect theory of Tversky and Kahneman (1992) suggests

$$v(F) = \int_{-\infty}^{\infty} u(x \mid x_0) \pi \left(f(x), \operatorname{sign}(x - x_0) \mid x_0 \right) dx \tag{5}$$

where x_0 is a reference point (usually assumed to be the status quo or current wealth), u is a function that is concave over outcomes that are greater than the reference point (representing risk aversion) and convex over outcomes that are less than the reference point (representing loss aversion), and π is a positive function that differs depending on whether $x - x_0$ is positive or negative.

We refer to all models that modify the utility and/or probability representations of the EU model as modified expected utility (MEU) models. Many of these models (e.g., Machina 1982, Chew 1983, Quiggin 1982) are implied by a set of behavioral axioms intended to ensure that the individual meets some rational criteria given the probability and outcome combinations available as possible choices. Others (e.g., prospect theory) are built on assumptions intended to mirror human behavior.

This class of models is not all-encompassing. Other models are based on similarity (e.g., Rubinstein 1988) or ad hoc properties of prospects (e.g., Lichtenstein and Slovic 1971). However, MEU models, and primarily prospect theory applications, make up the overwhelming bulk of non-expected utility models that have found practical empirical application in areas ranging from finance (e.g., Benartzi and Thaler 1995; Barbeis, Huang, and Santos 2001) to medical decisions (Verhoef, De Haan, and Van Daal 1994).

For empirical application, researchers have proceeded in two primary directions. Many have used data to estimate the underlying risky distributions, and then used these estimates as inputs to estimate or calibrate the parameters of the decision function (e.g., Barbeis, Huang, and Santos 2001). Alternative approaches have examined decision data for patterns that are consistent with a particular model without estimating parameters (e.g., Jullien and Salanie 2000). These two research methodologies are similar to the general literature employing EU as a decision model. For example, Chavas and Holt (1996) assume specific parametric forms for risk preferences and perceptions of the consequences of risky decisions, and then estimate the associated parameters to explain observed behavior. Alternatively, Chavas and Pope (1985) test certain behavioral restrictions in a reduced form production model based on Pope (1980) to determine some broad characteristics of risk aversion can be rejected in an EU framework. Elsewhere, we have shown that such estimates are unidentified without imposing severely restrictive functional forms on either the preferences or perceptions (Just and Just, forthcoming).

The Illusory Advantage of Experimental Methods for Risk Model Discernment

Early efforts to identify both preferences and perceptions jointly were based on field data. Experimental approaches have become more popular, particularly in the last decade. Experimentalists have argued that econometric tests using field data are not as powerful for excluding choice models as are experimental methods. The basis for this claim is that experimentalists can actually observe and manipulate probabilities and payoffs associated with the various prospects. The experimental literature has used two analogous strategies to test these theories. One approach cleverly manipulates the probability distributions to test for violations of models (e.g., MacCrimmon and Larsson 1979; Lichtenstein and Slovic 1971). Alternatively, some have used laboratory choices to estimate the choice function parameters and determine the best performing model (e.g., Hey and Orme 1994; Buschena and Zilberman 2009), or examine the overall fit (e.g., Tversky and Kahneman 1992).

Whether based on experimental or field data, however, each of these methods take the researcher's assessment of probability distributions as the input for the model tests and use this to identify either the probability distortion or the utility modification scheme. Our work calls this procedure into question and further questions the feasibility of discerning the proper class of models using *any* technique when modifications of both preferences and perceptions are contemplated.

Early work by Edwards (1955) discovered that individuals distort probability perceptions in some predictable (although erratic) ways. Within the context of each of the MEU models, the decision process occurs in the following stages:

(i) The individual is given information about various outcomes and their likelihoods.

(ii) The individual forms subjective perceptions about the likelihood of various possibilities that, through errors in processing information and possible affects of prior experience, may be represented by distorted probabilities.

(iii) The consequences of various possibilities are evaluated, as may be represented by a modified utility function, which may be affected by the perceived likelihood of possibilities or past experience.

(iv) A choice is made based upon the perceptions in (ii) and preferences in (iii).

In each case, the researcher is able to observe (i) and (iv), but not (ii) and (iii). Even when the experimentalist provides the exact probabilities to the subjects, these probabilities and the resulting perceptions formed in (ii) may not represent belief formation in more practical experiences. Probabilities may be interpreted very differently from experiential frequencies. For example, how does an untrained subject relate to a statement that something will occur 0.5% of the time or even 5 times out of a thousand if they have not faced those odds within the realm of practical experience, and can readily identify the occasions when they have faced such odds. There is still a process of interpretation (separate from distortion) that cannot be observed.

Those studying decision-making have often overcome this problem by imposing a restrictive functional form for either (ii) or (iii) or both. Similarly, the judgment bias literature has overcome this problem primarily by imposing a risk-neutral expected utility decision process. Because the intermediate steps of belief formation and valuation are not observed, the researcher cannot separately identify these two processes. In the next section, we show that each of the MEU models modifies at least one of these component functions in a way that is mathematically equivalent to potential modifications of the other component. Thus, if both perception and preference modifications are admitted, then neither are econometrically identified separately by observed choices. This is true of both field data and experimental data. That is, experimental analysis that tests for modification of the utility function relies heavily on the assumption that the researcher's perceptions of (i) are also the subjective perceptions of the decision-maker in (ii).

Inability to Discern Preference and Perception Modifications Based on Behavior

In this section we show that preference modifications cannot be discerned from perception modifications nor vice versa based on behavioral data. The results imply that preference and perception modifications are not econometrically identifiable in models that admit both. The implication is that identification of models that admit either one is possible only by arbitrarily excluding the other.

Generalized Expected Utility

Consider first the generalized expected utility model of Machina (1982). This model is general enough to include many other models as special cases (e.g., Edwards 1955; Chew 1983). To apply the model in (4) to consider choice under risk, suppose an agent must make a possibly vector-valued choice $\boldsymbol{q}$ from a set $Q \subset \Re^m$ that determines the specific distribution $F(x \mid \boldsymbol{q})$ from which the agent's payoffs are drawn. If $F(x \mid \boldsymbol{q})$ is a valid distribution then $\lim_{x\to-\infty} F(x \mid \boldsymbol{q}) = 0$, $\lim_{x\to\infty} F(x \mid \boldsymbol{q}) = 1$, and $dF(x \mid \boldsymbol{q}) \geq 0$. Because $\boldsymbol{q}$ determines F, Machina's generalized utility function $u(x \mid F)$ in (4) can be represented as $u(x \mid \boldsymbol{q})$, i.e., the only effect of the choice on F is through $\boldsymbol{q}$. The choice criterion under generalized expected utility is thus represented as

$$\max_{\boldsymbol{q}} v(\boldsymbol{q}) = \int_{-\infty}^{\infty} u(x \mid \boldsymbol{q}) dF(x \mid \boldsymbol{q}). \tag{6}$$

The following propositions consider the possibility of judgment bias represented by misperceptions of $F(x \mid \boldsymbol{q})$.

Assumption. Admissible modified utility functions of the form $u(x \mid \boldsymbol{q})$ are assumed to be bounded, positively monotonic, and, without loss of generality, positive for all $x \in \Re^1$ and $\boldsymbol{q} \in Q$. Concavity can be further assumed without altering the results.

Proposition 1 (*Failure of Model Identification*). Suppose an agent facing a choice set $\boldsymbol{q} \in Q$ possesses preferences described by $u(x \mid \boldsymbol{q})$ and perceptions described by the distribution $F(x \mid \boldsymbol{q})$. Then many specification pairs $\{\tilde{u}(x \mid \boldsymbol{q}), \tilde{F}(x \mid \boldsymbol{q})\}$ exist that generate behavior identical to the true specification pair $\{u(x \mid \boldsymbol{q}), F(x \mid \boldsymbol{q})\}$.

Proof: Consider an arbitrary set of preferences represented by $\tilde{u}(x\,|\,\boldsymbol{q})$. Then (mis)perceptions denoted by the distribution $\tilde{F}(x\,|\,\boldsymbol{q})$ and defined by

$$d\tilde{F}(x\,|\,\boldsymbol{q}) \equiv u(x\,|\,\boldsymbol{q})dF(x\,|\,\boldsymbol{q})/\tilde{u}(x\,|\,\boldsymbol{q})$$

yield the choice criterion

$$\max_{\boldsymbol{q}} v(\boldsymbol{q}) = \int_{-\infty}^{\infty} \tilde{u}(x\,|\,\boldsymbol{q})d\tilde{F}(x\,|\,\boldsymbol{q}) \equiv \int_{-\infty}^{\infty} u(x\,|\,\boldsymbol{q})dF(x\,|\,\boldsymbol{q}),$$

which is mathematically identical to the true choice criterion. Thus, observed behavior alone cannot discern which of the two generalized utility functions represents preferences when the true (mis)perceptions of the agent are unknown.

Because the choice of alternative preferences in this proof is represented by an arbitrary function of x given $\boldsymbol{q}$, this proof implies that an infinite set of preferences (including dramatically different preferences) can fit behavior when the possibility of misperceptions associated with judgment bias is admitted. Given positivity and boundedness of u and $\tilde{u}$, $\lim_{x\to-\infty} \tilde{F}(x\,|\,\boldsymbol{q}) = 0$ is assured by $\lim_{x\to-\infty} F(x\,|\,\boldsymbol{q}) = 0$, and $d\tilde{F}(x\,|\,\boldsymbol{q}) \geq 0$ is assured by $dF(x\,|\,\boldsymbol{q}) \geq 0$. However, $\lim_{x\to\infty} \tilde{F}(x\,|\,\boldsymbol{q}) = 1$ is implied by $\lim_{x\to\infty} F(x\,|\,\boldsymbol{q}) = 1$ only with an additional condition that $\lim_{x\to\infty} u(x\,|\,\boldsymbol{q})/\tilde{u}(x\,|\,\boldsymbol{q}) = 1$. We do not require the latter condition in Proposition 1 because some modified expected utility theories admit judgment bias whereby distributions do not have unit measure over their domain (e.g., Preston and Baratta 1948). If such a restriction is imposed, then the latter condition must be added to define the arbitrary choice set for $\tilde{u}(x\,|\,\boldsymbol{q})$, but clearly an infinite set of such possibilities with dramatically different implications for risk behavior remains.

Proposition 2 (*Necessity of Arbitrary Assumptions*). Let U be a set of possible modified utility specifications $\tilde{u}(x\,|\,\boldsymbol{q})$ and Ψ be a corresponding set of possible perceptions $\tilde{F}(x\,|\,\boldsymbol{q})$ corresponding to any subset of the indistinguishable pairs of specifications associated with $\{u(x\,|\,\boldsymbol{q}), F(x\,|\,\boldsymbol{q})\}$ in Proposition 1. If a flexible specification of the modified utility function that admits all choices in U can be identified econometrically, then identification is accomplished by imposing an arbitrary assumption restricting misperceptions to a single possibility $\tilde{F}(x\,|\,\boldsymbol{q}) \in \Psi$.

Proof: From the proof of Proposition 1, for each $\tilde{u}(x \mid \boldsymbol{q}) \in U$ a corresponding $\tilde{F}(x \mid \boldsymbol{q}) \in \Psi$ exists such that observed behavior is explained identically by $\{\tilde{u}(x \mid \boldsymbol{q}), \tilde{F}(x \mid \boldsymbol{q})\}$. Thus, a unique $\tilde{u}(x \mid \boldsymbol{q}) \in U$ is identified only by imposing a unique $\tilde{F}(x \mid \boldsymbol{q}) \in \Psi$.

Proposition 2 implies that if the incorrect specification of (mis)perceptions is imposed, then the possibility of identifying the correct specification of the modified utility function may be eliminated by assumption. However, unlike many misspecification problems in econometrics where more flexible parametric specifications facilitate better approximation, in this case the estimated form may not reflect the correct risk preferences even approximately unless the misspecification of perceptions is approximately accurate. The same problem exists in reverse. That is, risk preferences are identified only by imposing an arbitrary restriction on the modified utility function. Thus, observing behavior alone cannot guarantee approximation in the estimation of either preferences or perceptions.

Proposition 3 (*Inability to Discern Models Based on Behavior*). Under the assumptions of Proposition 2, if flexible specifications of the modified utility function and perceptions admitting all choices in U and Ψ, respectively, can be found, then neither is econometrically identified by estimation of behavioral equations.

Proof: If many pairs of specifications $\{\tilde{u}(x \mid \boldsymbol{q}), \tilde{F}(x \mid \boldsymbol{q})\}$ in $U \times \Psi$ generate mathematically identical choice criteria and behavior, then parameters that select a specific modified expected utility function $u(x \mid \boldsymbol{q})$ from U and a specific representation of (mis)perceptions $F(x \mid \boldsymbol{q})$ from Ψ are not identified because the estimable functions are mathematically identical and thus provide the same fit of observed behavior.

Proposition 3 suggests a futility associated with the continued effort to (a) identify the correct representation of a modified utility function that can represent behavior under the potentially false assumption of perfect perceptions, (b) identify a standard representation of misperceptions that can explain choice under risk ignoring the potential for modification of the utility function, or (c) attempts to identify both modifications of the utility function and misperceptions of the risk distribution simultaneously. Certainly, when based on revealed preferences in behavioral data alone, none of these approaches are valid if both MEU and judgment bias are admitted. While real-world behavioral data may be supplemented by experimental approaches in an attempt to provide further identification, such approaches suffer from these same problems when the failure of ability to observe steps (ii) and (iii) in the previous section is recognized.

Rank-Dependent Expected Utility

Consider next the rank-dependent utility model of Quiggin (1982). To operationalize his model for the case of choice under risk, suppose an agent must make a possibly vector-valued choice $\boldsymbol{q}$ from a set $Q \subset \Re^m$ that determines the specific distribution $p(\boldsymbol{q}) \equiv \{p_1(\boldsymbol{q}),\ldots, p_n(\boldsymbol{q})\}$ from which the agent's payoffs are drawn. We assume, as in typical operationalizations of Quiggin's model, that the possible states, $x_1,\ldots,x_n$, and thus their ranking, are fixed (see for example, Tversky and Kahneman 1992). This assumption is innocuous when some states can have zero probability under alternative choices. Then the decision criterion associated with (1) becomes

$$\max_{\boldsymbol{q}} v(p(\boldsymbol{q})) = \sum_{i=1}^{n} x_i \pi_i(p(\boldsymbol{q}))$$

where the weighting function follows (2). If these weights sum to one, i.e., $\pi_1(p(\boldsymbol{q})) + \ldots + \pi_n(p(\boldsymbol{q})) = 1$, then this model can be trivially represented alternatively as maximization of standard von Neumann-Morgenstern utility under risk neutrality where the probabilities of outcomes are misperceived as $\tilde{p}_i(\boldsymbol{q}) \equiv \pi_i(p(\boldsymbol{q})),\ i = 1,\ldots,n.$ If the weights do not sum to one, then the criterion can be written as

$$\max_{\boldsymbol{q}} v(p(\boldsymbol{q})) = \sum_{i=1}^{n} u(x_i \mid \boldsymbol{q}) \tilde{p}_i(\boldsymbol{q})$$

where

$$\tilde{p}_i(\boldsymbol{q}) \equiv \pi_i(p(\boldsymbol{q})) \Big/ \sum_{i=1}^{n} \pi_i(p(\boldsymbol{q})),$$

$$u(x_i \mid \boldsymbol{q}) = x_i \sum_{i=1}^{n} \pi_i(p(\boldsymbol{q})),\ i = 1,\ldots,n.$$

Once the model is in this form, the proof for the generalized utility model of Machina becomes applicable, yielding the same conclusions as in Propositions 1-3.

Prospect Theory

To consider prospect theory, we use the modifications introduced by Tversky and Kahneman (1992), which are applied in a manner similar to Quiggin's rank-dependent utility model. The difference is that a $u(x_i)$ function is introduced in place of x_i in the value function, in which case Quiggin's value function becomes

$$\max_{\boldsymbol{q}} v(p(\boldsymbol{q})) = \sum_{i=1}^{n} u(x_i)\pi_i(p(\boldsymbol{q}))$$

Assuming the reference point x_0 is not endogenous to the choice, which is the standard approach in the literature, this function can represent a discrete version of the value function in (5) after conditioning both the utility and probability weights on the reference point and conditioning the distribution on the choice variable,

$$v(\boldsymbol{q}) = \sum_{i=1}^{n} u(x \mid x_0)\pi\left(f(x \mid \boldsymbol{q}), \text{sign}(x - x_0) \mid x_0\right).$$

That is, the two sets of probability weights (one for gains and one for losses) can be adequately reflected in this framework after conditioning probability weights on the reference point. In this case, the proofs of Propositions 1-3 hold piecewise above and below the reference point (which is the reason for assuming the reference point is exogenous). Thus, the results of Propositions 1-3 apply.

For practical purposes, even though the reference point is exogenous, it can be potentially identified if observations are made with different reference points. Hence, we focus only on the properties of the value function and the probability weight given a reference point.

General Implications

Because the choice of alternative risk preferences represented by $\tilde{u}(x \mid \boldsymbol{q})$ in the proof of Proposition 1 is arbitrary, the equivalence that generates Propositions 1-3 holds for an infinite set of pairs of $\{\tilde{u}(x \mid \boldsymbol{q}), \tilde{F}(x \mid \boldsymbol{q})\}$. Thus, Propositions 1-3 imply that no possibilities exist for validating or rejecting the Machina model from revealed preference (behavioral) data once errors in perceptions are admitted. Because the cases of Quiggin's rank-dependent utility and prospect theory can be generally reduced to the same framework, the same conclusion applies to those general alternatives to expected utility theory as well.

Can the Expected Utility Hypothesis Be Rejected?

A further application of the approach of this paper raises serious questions about the context of results that claim to invalidate the standard von Neumann-Morgenstern expected utility hypothesis.

Proposition 4 (*Inability to Reject the Expected Utility Hypothesis*). Suppose econometric analysis of behavioral data estimates a model denoted by $\{u(x \mid \boldsymbol{q}), F(x \mid \boldsymbol{q})\}$. Then a set of (mis)perceptions exists that explains the data equally well under the von Neumann-Morgenstern expected utility hypothesis.

Proof: Suppose in reality the agent has a standard bounded nonnegative von Neumann-Morgenstern utility function $\tilde{u}(x)$ representing arbitrary risk preferences, i.e., is independent of the distribution of payoffs represented by $\boldsymbol{q}$. But suppose judgment bias affects the agent's perceptions of the distribution of payoffs $F(x \mid \boldsymbol{q})$ and its dependence on the choice of $\boldsymbol{q}$ such that the perceived distribution is $\tilde{F}(x \mid \boldsymbol{q})$ defined by

$$d\tilde{F}(x \mid \boldsymbol{q}) \equiv u(x \mid \boldsymbol{q}) dF(x \mid \boldsymbol{q}) / \tilde{u}(x). \tag{7}$$

Then the choice criterion is

$$\max_{\boldsymbol{q}} v(\boldsymbol{q}) = \int_{-\infty}^{\infty} \tilde{u}(x) d\tilde{F}(x \mid \boldsymbol{q}) \equiv \int_{-\infty}^{\infty} u(x \mid \boldsymbol{q}) dF(x \mid \boldsymbol{q}),$$

which is mathematically identical and thus fits observed behavioral data exactly as well as the MEU model assuming MEU represented by $u(x \mid \boldsymbol{q})$ and perceptions represented by $F(x \mid \boldsymbol{q})$.

Proposition 4 shows that Machina's generalized expected utility model under his assumption that rejects errors in perception cannot achieve empirical superiority over the standard expected utility model after allowing misperceptions. In contrast, Machina assumes absence of misperceptions of probabilities. Although he uses this assumption to avoid failure of monotonicity, such an assumption is a high and unrealistic price to pay for attaining monotonicity given ample evidence of judgment bias (see below).

The proof of Proposition 4 allows the possibility that the misperceived distribution does not have unit measure on the domain of payoffs, as is allowed with some probability weighting models (as noted above). As in Proposition 1, with positivity and boundedness of u and $\tilde{u}$, $\lim_{x\to-\infty}\tilde{F}(x\,|\,\boldsymbol{q})=0$ is assured by $\lim_{x\to-\infty}F(x\,|\,\boldsymbol{q})=0$, and $d\tilde{F}(x\,|\,\boldsymbol{q})\geq 0$ is assured by $dF(x\,|\,\boldsymbol{q})\geq 0$, but $\lim_{x\to\infty}\tilde{F}(x\,|\,\boldsymbol{q})=1$ is implied by $\lim_{x\to\infty}F(x\,|\,\boldsymbol{q})=1$ only with an additional condition that $\lim_{x\to\infty}u(x\,|\,\boldsymbol{q})/\tilde{u}(x)=1$.

The latter condition may appear to be a serious limitation. However, all it requires is that $\boldsymbol{q}$ affects the valuation of particular outcomes in a way that does not introduce a scaling over the domain of outcomes. For example, Chew's (1983) special case of Machina's model in (3) avoids scaling by a normalization of utility weightings whereby

$$u(x\,|\,F)=a(x)u(x)\Big/\int_{-\infty}^{\infty}a(x)dF(x).$$

This modified utility function under correct perceptions can obviously be expressed equivalently with a standard von Neumann-Morgenstern utility function $u(x)$ and misperceptions represented by

$$d\tilde{F}(x)=a(x)dF(x)\Big/\int_{-\infty}^{\infty}a(x)dF(x),$$

which satisfies $\lim_{x\to\infty}\tilde{F}(x\,|\,\boldsymbol{q})=1.$ In addition, typical practical applications of Machina's generalized expected utility model (e.g., Machina 1989 and 1995) involve revising the utility evaluated at a particular payoff x only according to properties of the distribution at that point. Thus, the modified utility function in (4) yields $\lim_{x\to\infty}u(x\,|\,\boldsymbol{q})$ $\equiv\lim_{x\to\infty}u(x\,|\,F(x\,|\,\boldsymbol{q}))=\lim_{x\to\infty}u(x\,|\,1),$ which is independent of $\boldsymbol{q}$ and thus implies that $\lim_{x\to\infty}u(x\,|\,\boldsymbol{q})/\tilde{u}(x)=1$ is an innocuous requirement (given irrelevance of affine transformations).

Proposition 4 implies that anomalous behavior cannot reject the standard expected utility model when judgment bias is admitted. That is, observed behavioral data alone cannot distinguish one model from the other. The ex post evidence can favor the generalized expected utility model with correct perceptions no more or less than the standard von Neumann-Morgenstern utility model under misperceptions. Any test comparing the performance of the standard von-Neuman-Morgenstern utility model to the generalized expected utility model must rely on dubious assumptions that individuals have perfect perceptions. Without this identifying assumption, the evidence cannot possibly reject one model in favor of the other because they imply identical behavior.

Proposition 5 (*Failure of Identification of Arrow-Pratt Risk Aversion Measures*). Suppose econometric analysis of behavioral data estimates a model denoted by $\{u(x), F(x \mid \boldsymbol{q})\}$. Then a set of (mis)perceptions exists that explains the data equally well under an arbitrary alternative set of preferences associated with different Arrow-Pratt measures of risk aversion.

Proof: The proof of Proposition 4 applies upon replacing $u(x \mid \boldsymbol{q})$ by $u(x)$ because the choice of $u(x)$ is arbitrary, which means the associated Arrow-Pratt measures of risk aversion are arbitrary. See Just and Just (forthcoming) for a similar proof.

Although the standard Arrow-Pratt concepts of absolute and relative risk aversion have proven highly useful for theoretical and conceptual work, Proposition 5 raises serious concerns about their empirical usefulness. If they are unidentifiable except by arbitrary parametric assumptions and arbitrarily ignoring judgment bias, perhaps some parallel but alternative concepts may be more useful for empirical purposes. For example, more useful empirical concepts of risk aversion would characterize behavioral relationships independent of unverifiable underlying assumptions. We discuss this possibility below.

Toward a Practical Methodology

Faced with the lack of econometric identification except through arbitrary assumptions, what is an empirical risk researcher to do? The results in this paper suggest three alternatives:

1. Continue with a futile research agenda based on behavioral data, ignoring the identification problem associated with anomalous behavior and the related unobservability problem of experimental methods.
2. Develop additional data as needed to identify and estimate the full structure of risky choice problems, i.e., achieve identification without grossly simplifying arbitrary assumptions.
3. Develop a practical approach for the typical research problem (which is limited to behavioral data) that recognizes possible anomalous phenomena and the related dilemma presented by the identification problem.

The Futile Agenda

If perception failures cannot be discerned from utility modifications based on behavioral data alone except with arbitrary assumptions, then the course of the existing risk research agenda is futile. Conflicting empirical results are to be expected. Dramatic differences in estimated preferences will occur due to arbitrary assumptions that attribute anomalous behavior to either utility modification or misperceptions, or to some specific combination of the two, as well as to specific parametric specifications representing either. Such a research agenda makes little sense because no hope exists for convergence of knowledge, of empirical results, or related policy implications. For example, what is the sensibility of estimating the generalized expected utility model or one of its special cases when judgment bias is ignored, given that judgment bias is well-documented in a now extensive literature.[1] More generally, given that the explanatory power of a continuum of models ranging from no utility modification to no judgment bias have identical explanatory power, econometric identification must be recognized as impossible except artificially when based on behavioral data alone. This is true whether data are naturally-occurring data from the real world or generated experimentally either in the lab or the field.

A Full-Structure Approach

Another approach is to attempt estimation of the full structure of the problem of risky choice behavior. Given the limitations of behavioral data clarified above, this requires generating additional data that can identify one of the individual structural components. We refer to the additional required data as non-behavioral data. For example, Kőzegi and Rabin (2007) propose elicitation of non-objective perceptions from the individuals generating the revealed-preference behavioral data in order to model revealed mistakes. Such an approach was used by Just, Calvin and Quiggin (1999) to separate preferences from perceptions in order to decompose the incentive for crop insurance participation into risk aversion and asymmetric information components. However, such an approach requires an intense data-gathering effort because the elicited survey observations representing perceptions must be paired with the revealed-preference

[1] Our critique also applies equally to the judgment bias literature, which relies on severe assumptions regarding preferences or communication to identify perceptions. However, time-based effects such as recency or primacy (Hogarth and Einhorn 1992) at least seem to establish that misperceptions play some role—given that the order of observation does not affect preferences.

(behavioral) observations generated by the same individuals. This approach may be impractical for many problems. Certainly, such data are unlikely to be available for many policy analyses at a level that is representative of a full population.

When this approach is applied to problems of production or household transformation problems, estimation of the full structure of the risk problem also requires estimation of the technology and its risk implications. This may require estimation of stochastic perceptions for both prices and technologies as affected by random production conditions, e.g., as in agriculture.

Interestingly, typical production estimation methods in the 1960s involved joint estimation of behavior, represented by first-order conditions, and the production technology, represented by a production function (Mundlak and Hoch 1965; Zellner, Kmenta and Dreze 1966). This permitted estimation of both the deterministic and stochastic implications of technology based on physical input-output data that was generally paired with data used to estimate behavioral equations. If individuals form objective perceptions based on actual technological performance, then this approach can solve part of the structural estimation problem.

In contrast, applications of duality have led to estimation of derivatives of profit and cost functions assuming risk neutrality. While the technologies implied by estimated specifications were recoverable, they generally were not used to test or validate the implicit technological relationships. Thus, such practices were a case where assumptions were substituted for identification, imposing the rather extreme preference structure of risk neutrality. Estimation of the full structure of a risk problem would require adding this structure and related perception errors to the model explaining risk behavior in order to estimate preferences with a full structural approach (see Just and Just, forthcoming).

While supplementing the estimation of behavioral equations with estimation of technological equations may call for resurrecting some aspects of primal approaches that preceded duality, estimation of perceptions of other stochastic phenomena by means of elicitation may be more controversial. Imposing objective models of perception formation may ignore misperceptions. After all, not even researchers have generally agreed on the best model to explain expectations (rational verses ARCH/GARCH versus Bayesian models). Furthermore, attempts to determine which parametric distributional form explains behavioral data has led to a great deal of controversy and lack of identification. For example, Anderson et al. (2001) have found that stock returns appear to be log-normally distributed in contrast to accepted wisdom that has assumed normality. Just and Weninger (1999) highlight the controversy associated with identification of the correct distribution of crop yields for crop insurance purposes, where correct modeling of the tails of the distribution is critical. Finally, elicitation approaches are subject to the traditional criticisms of hypothetical, strategic, and/or temporal bias (Whitehead and Blomquist 2006).

Thus, the full structural approach presents a challenging data collection problem as well as a complex structural estimation problem. Rich policy implications are facilitated by uncovering the full structure of risky choice problems, but the expense of data collection and the complexity of estimation problems are likely to make this approach impractical except in relatively few problems (for example, it may be feasible in an experimental laboratory). Furthermore, the vulnerability to various sources of bias is likely to raise a host of potential criticisms and obstacles to clear and robust identification.

A Practical Approach

In contrast to full structural estimation, we define the practical case as the case where only behavioral data are available. Recognizing that separate identification of preferences and perceptions is impossible in this case (except by arbitrary assumptions), a sensible approach is to focus on estimation of the value function $v(F)$ directly. This avoids the inability to identify multiplicative functions involving common variables. While direct estimation of $v(F)$ may seem impractical because theory offers limited guidance for specification without an underlying decision framework, the following proposition is helpful.

Proposition 6 (*A Universal Linear Mean-Variance Specification*). Where $v(F(\boldsymbol{q}))$ is the value function that orders alternative risky choices given the choice set $\boldsymbol{q} \in Q$, suppose the distribution F is characterized by its mean μ, variance σ, and any other necessary identifying parameters in θ so the value function can be represented as $v(\mu,\sigma,\theta)$. Assuming monotonicity ($v_\mu > 0$), this choice criterion can be represented in a linear mean-variance form $\mu(\boldsymbol{q}) - (\varphi/2)\sigma(\boldsymbol{q})$ for some $\varphi = \varphi(\mu,\sigma,\theta)$. However, this φ does not generally measure absolute risk aversion, nor is the associated first-order condition generally of the form $\mu_{\boldsymbol{q}}(\boldsymbol{q}) - (\varphi/2)\sigma_{\boldsymbol{q}}(\boldsymbol{q})$.

Proof: Without loss of generality, because a positive monotonic transformation of $v(\cdot)$ preserves the ordering, let $\mu = v(\mu,0)$. Then $v(\mu,\sigma) = \mu(\boldsymbol{q}) - (\varphi/2)\sigma(\boldsymbol{q})$ is always satisfied by defining $\varphi(\mu,\sigma,\theta) = 2[\mu(\boldsymbol{q}) - v(\mu,\sigma,\theta)]/\sigma(\boldsymbol{q})$. The first-order condition for this criterion is $v_{\boldsymbol{q}} = \mu_{\boldsymbol{q}}(\boldsymbol{q}) - (\varphi_{\boldsymbol{q}}/2)\sigma(\boldsymbol{q}) - (\varphi/2)\sigma_{\boldsymbol{q}}(\boldsymbol{q})$, which adds an additional term compared to the common form stated in the proposition. To prove by exception that φ does not measure absolute risk aversion generally, solving the differential equation $\varphi = -u''/u'$ that defines CARA implies that the utility function, aside from affine transformations, can be represented as $u(x) = -e^{-\varphi x}$. Assuming any two-parameter distribution with a closed-form expression for its moment generating function, e.g., a uniform distribution on $(\mu - \sqrt{3\sigma}, \mu + \sqrt{3\sigma})$, yields an expression for $v(\mu,\sigma)$ that permits derivation of an analytic expression for the risk premium. The resulting risk premium will not have the form $(\varphi/2)\sigma$ except under normality. The tedious derivation for this case is omitted for convenience and brevity.

Proposition 6 is trivial but included here because some published studies ignore these facts.[2]

Given this result, we suggest applying the principle of Occam's razor, which is to use the simplest and most well-understood model when no distinction is possible. The value function suggested by Proposition 6 consists of two terms, the mean function $\mu(\boldsymbol{q})$ and the risk premium defined by $r(\mu,\sigma,\theta) \equiv (\varphi/2)\sigma(\boldsymbol{q})$. When combined, these yield the certainty equivalent $c(\mu,\sigma,\theta) \equiv \mu(\boldsymbol{q}) - (\varphi/2)\sigma(\boldsymbol{q})$. The mean, risk premium, and certainty equivalent are singularly well-understood terms and are exceptionally intuitively appealing. Further, because they have additive relationships, they can be identified increasingly accurately by using increasingly flexible functional forms.

Furthermore, these forms suggest a useful characterization of risk aversion measurement for empirical purposes. That is, because the Arrow-Pratt measures of risk

---

[2] This type of model is reminiscent of Meyer's (1987) location and scale condition that permits certain types of two-moment distributions to be represented by a mean-standard-deviation model. However, the condition here is general whereas Meyer's model requires that the payoff of interest is a linear function of the mean and standard deviation. His condition excludes, for example, the CRRA case with log-normality.

aversion are unidentifiable by Proposition 5 in the practical case of data availability (except by arbitrary parametric assumptions), more intuitively appealing and truly identifiable concepts of risk aversion can be tested if the theoretical concepts are generalized for empirical purposes. For example, a somewhat more general concept of constant absolute risk aversion (CARA) that can be tested in this model is represented by $r(\mu,\sigma,\theta) = a\sigma$ for some constant a. In the special case of normality, this corresponds to the true Arrow-Pratt CARA concept. However, imposing normality on perceptions is not sensible because it cannot be tested in the practical case even though imposing it can lead to severe errors in characterizing preferences. This empirical version of CARA offers a more honest empirical representation of the concept, because virtually all studies that claim to estimate CARA models simply impose an assumption of normality.

Further, this approach encourages better model validation. For example, few empirical studies that assume an Arrow-Pratt CARA model attempt to validate it against a more general parametric form of $r(\mu,\sigma,\theta)$ because convenient forms for broader maintained hypotheses have not been popularized. Our approach suggests ready alternative hypotheses.

Turning to an empirical version of constant relative risk aversion (CRRA), a truly identifiable hypothesis that can be tested is given by $r(\mu,\sigma,\theta) = \mu s(\sigma/\mu^2)$ where $s(\cdot)$ is a positive monotonic function. This makes the risk premium proportional to the mean where relative risk is determined by the coefficient of variation. In the special case of Arrow-Pratt CRRA under log-normality, the risk premium is of the form $r(\mu,\sigma,\theta) = \mu[1 - (1 + \sigma/\mu^2)^{-\psi/2}]$ where ψ is the Arrow-Pratt measure of CRRA.[3] Thus, this hypothesis corresponds to the true Arrow-Pratt CRRA concept under log-normality. However, imposing log-normality on perceptions is not sensible because it cannot be tested in the practical case even though

[3] Where $\ln x \sim N(\mu^*,\sigma^*)$, the normal moment generating function implies $\mu = E(e^{\ln x}) = M_{\ln x}(1) = e^{\mu^* + \sigma^*/2}$ and $E(x^2) = E(e^{2\ln x}) = M_{\ln x}(2) = e^{2\mu^* + 2\sigma^*}$. Solving these relationships where $\sigma = E(x^2) - \mu^2 = e^{2\mu^* + \sigma^*}(e^{\sigma^*} - 1)$ obtains $\mu^* = \ln\mu - \ln(1 + \sigma/\mu^2)/2$ and $\sigma^* = \ln(1 + \sigma/\mu^2)$. Solving the differential equation $\psi = -u''x/u'$ that defines CRRA implies that the utility function, aside from affine transformations, can be represented as $u(x) = (1 - \psi)^{-1}x^{1-\psi} = (1 - \psi)^{-1}e^{(1-\psi)\ln x}$ where $u'' < 0$ implies $\psi > 0$. The moment generating function associated with $\ln x \sim N(\mu^*,\sigma^*)$ yields $Eu(x) = (1 - \psi)^{-1}E(e^{(1-\psi)\ln x}) = (1 - \psi)^{-1}e^{(1-\psi)[\mu^* + (1-\psi)\sigma^*/2]}$, which implies that the decision criterion under CRRA and log-normality is $u^{-1}(Eu(x)) = \mu^* - (\psi - 1)\sigma^*/2 = \ln\mu - (\psi/2)\ln(1 + \sigma/\mu^2)$. The certainty equivalent, which approaches μ as σ approaches zero, is thus $e^{\ln\mu-(\psi/2)\ln(1+\sigma/\mu^2)} = \mu(1 + \sigma/\mu^2)^{-\psi/2} = \mu - r(\mu,\sigma)$ where the risk premium is $r(\mu,\sigma) = \mu[1 - (1 + \sigma/\mu^2)^{-\psi/2}]$.

imposing it can lead to severe mischaracterization of preferences. Again, this approach seems to be a more honest representation of empirical results that encourages better model validation.

An Example

To illustrate this approach in the case where the distribution of outcomes is completely characterized by the mean and variance (or any two parameters that distinctly determine mean and variance), suppose the parametric specification of the risk premium is $r(\mu,\sigma) = A\mu^{\alpha}\sigma^{\beta}$. Then the first-order condition for maximization of $v(\mu,\sigma) = \mu - r(\mu,\sigma)$ is $v_{\boldsymbol{q}} = \mu_{\boldsymbol{q}} - r_{\mu}\mu_{\boldsymbol{q}} + r_{\sigma}\sigma_{\boldsymbol{q}} = 0$ becomes

$$\mu_{\boldsymbol{q}} = \alpha A\mu^{\alpha-1}\sigma^{\beta}\mu_{\boldsymbol{q}} + \beta A\mu^{\alpha}\sigma^{\beta-1}\sigma_{\boldsymbol{q}} = (\alpha\mu_{\boldsymbol{q}}/\mu + \beta\sigma_{\boldsymbol{q}}/\sigma)r. \tag{8}$$

Thus, simple and intuitively plausible parametric specifications for $\mu(\boldsymbol{q})$ and $\sigma(\boldsymbol{q})$ can permit empirical implementation. For example, equation (8) can be estimated directly by GMM methods. Alternatively, where $\boldsymbol{q}$ is a vector with elements q_i and q_j, taking ratios reveals a relatively simple form,

$$\frac{\mu_{q_i}}{\mu_{q_j}} = \frac{\alpha\mu_{q_i}/\mu + \beta\sigma_{q_i}/\sigma}{\alpha\mu_{q_j}/\mu + \beta\sigma_{q_j}/\sigma}.$$

This approach offers several empirical conveniences. First, the standard risk properties $v_{\mu} > 0$ and $v_{\sigma} < 0$ are exceptionally easy to test or impose in estimation. Second, tests for special cases of risk aversion are exceedingly simple. For example, empirical CARA as defined above simply corresponds to $\alpha = 0$, $\beta = 1$. Decreasing absolute risk aversion is represented by $\alpha < 0$. Empirical CRRA as defined above simply corresponds to $\alpha = 1 - 2\beta$. Empirical decreasing relative risk aversion is represented by $\alpha < 1 - 2\beta$. Third, improving the accuracy of approximation by adding functional flexibility is straightforward. Plausible forms for $r(\mu,\sigma)$ might include a translog or quadratic form in μ and σ. Also, additional parameters reflecting skewness or downside risk can be added to the risk premium specification to the extent data are sufficient for identification.

Accommodating Judgment Bias

The convenient and intellectually honest aspect of this approach is that most common forms of judgment bias are easily accommodated while clearly reflecting the lack of econometric identification. Overconfidence is a typical form of judgment bias whereby the distribution is treated as narrower than it is in reality (Alpert and Raiffa 1982; Camerer 1995; Lichtenstein 1982; Murphy 1974). Overconfidence can be reflected by a simple parameter that reduces the role of σ in the risk premium, e.g., be replacing σ with σ^b where b is an unknown parameter in the unit interval. In the particular specification above, it is immediately apparent that b and β cannot be identified separately in the practical case where only behavioral data are available. In $r(\mu,\sigma) = A\mu^{\alpha}\sigma^{b\beta}$, the combined parameter $b\beta$ can be estimated but lower risk aversion cannot be differentiated from over simplification. The implication is that the level of risk aversion attributed empirically to economic behavior in typical risk studies may, in part, incorrectly represent judgment bias in the form of oversimplification. This is a possible explanation for the results of Just and Peterson, who find that the amount of estimated risk response in agriculture exceeds the theoretical limits that can be attributed to risk aversion.

Other forms of judgment bias associated with dynamic behavior, such as representativeness and conservatism, can be incorporated by modeling σ as a weighted average, $\tau\sigma_r + (1 - \tau)\sigma_d$, of recent variation σ_r and variation in the more distant past σ_d. Representative bias is a tendency to overweight newly observed outcomes and underweight prior or old information, an explanation suggested for the overreaction of stock markets (DeBondt and Thaler 1985). Conservatism bias occurs when an individual weighs the status quo too heavily, which has been offered as an explanation for positive autocorrelation in forecast errors that are inconsistent with rational expectations (Edwards 1982). Both could be tested for practical risky choice problems with this specification.

The law of small numbers is a form of judgment bias that refers to the tendency to rely too heavily on experience when only a few observations are available (Tversky and Kahneman 1982). Law-of-small-numbers bias can be tested in this type of model by applying an adjustment factor to the sample variance that reflects a downward bias when the number of observations is small.

Accommodating Anomalous Behavior

Turning to representations of anomalous behavior, properties of prospect theory can also be easily accommodated. For example, the role of a reference point can be included by replacing the variance σ by the mean-squared error. Replacing $\sigma = E[(x - \mu)^2]$ by $E[(x - x_0)^2]$ where x_0 is the relevant exogenous reference point, the error squares identity implies $E[(x - x_0)^2] = E[(x - \mu)^2] + (x_0 - \mu)^2 \equiv \sigma + a$. Thus, a simple additive shifter of the variance, $\sigma + a$, facilitates estimation and testing of the role of the reference point x_0 where $a = (x_0 - \mu)^2$. This approach can be used to test whether the influence of a change in wealth on risk aversion is the same as an equal increase in the mean payoff. Other behavioral differences in response to downside versus upside risk can be accommodated by separating $\sigma = E[(x - \mu)^2]$ into $\sigma = \sigma_- + \sigma_+$ where $\sigma_- = E[(x - \mu)^2 \mid x < \mu]\Pr(x < \mu)$ and $\sigma_+ = E[(x - \mu)^2 \mid x \geq \mu]\Pr(x \geq \mu)$.[4] Any tendencies to overlook certain kinds of outcomes in assessing risk can be accommodated in the parametric specifications for μ and σ.

With this approach, the predominance of anomalous behavior can be tested in broader decision-making environments (non-experimental settings) that represent the bulk of economic behavior, at least to the extent identification is possible in the practical case where only behavioral data are available. Thus, the relevance of various factors for broadly applicable policy analysis can be clarified.

Policy and Welfare Analysis

As these potential representations of judgment bias and anomalous behavior are incorporated into the specifications for μ, σ, and r, the natural consequence is that identification of distinct parameters is weakened and perhaps undermined completely. However, the usefulness for policy and economic welfare analysis can be strengthened in the process. Replicable anomalies have been broadly regarded as an invalidation of both standard expected utility theory and of welfare economics (Ariely, Lowenstein, and Prelec 2003). However, recent developments show that welfare economics and policy analysis can be valid in spite of anomalies and judgment bias. Bernheim and Rangel (forthcoming) show how accounting for ancillary conditions can create a framework for welfare analysis that allows for anomalous behavior. Further work by Kozegi and Rabin (2008) derives a theory of revealed mistakes that can enhance the ability of welfare theory to address problems of self control or misperception. These theories provide promise for policy analysis in many cases where behavioral anomalies are common (e.g., retirement savings

[4] These approaches to incorporating behavior aspects that explain various forms of anomalous behavior may be impossible to differentiate from sensitivity of the risk premium to skewness or other higher moments of the distribution in absence of anomalous behavior, as rightfully they should. According to the results in this paper, an arbitrary utility function can fit any set of behavioral data if sufficient misperceptions are allowed. The practical approach suggested here does not eliminate these empirical ambiguities by arbitrary assumptions, but rather represents them explicitly for empirical analysis.

or investment) and make a clear argument for maintaining a well defined welfare methodology despite systematic errors in human choice.[5]

Recognizing errors in perceptions is more sensible than abandoning policy or welfare analysis. The concepts of willingness to pay (WTP), upon which modern welfare economics is founded, are well-defined with this approach even though econometric analysis may not be able to attribute some observed behavior distinctly to preferences versus perceptions. That is, WTP depends on perceptions as well as preferences.

Expanding the empirical basis for economic welfare analysis in this way enhances the ability to evaluate policies that offer information and training. Further, following the approach developed by Foster and Just (1989), the cost of ignorance, delay, and failure to disclose information can be readily evaluated using the full array of Hicksian welfare concepts.

Conclusions

Results show that innumerable "as if" models can fit behavioral data under risk when both modified expected utility and misperceptions (judgment bias) are admitted. In particular, when judgment bias is admitted, we show that behavioral data alone cannot reject the standard von Neumann-Morgenstern expected utility hypothesis. Thus, any statistical significance that suggests identification in the practical case where only behavioral data are available is the artifact of arbitrarily imposing a specific parametric form. A further implication is that arbitrary parametric specifications chosen for empirical modeling determine arbitrarily the Arrow-Pratt measures of risk aversion that are estimated. Thus, the standard Arrow-Pratt measures are not truly identified econometrically in the practical case where only behavioral data are observed.

These results exacerbate a problem we have previously identified whereby both preferences and perceptions cannot be separately identified from behavioral data in standard risk problems. The underlying problem is that separate econometric identification is not possible when two functions appear in multiplicative form and include common variables. Imposing a narrow parametric specification on either function can cause gross

[5] We also note that our suggested approach is an application of the principle of sufficient statistics illuminated by Chetty (2008) for the case where estimation may be able to reflect the implications of multiple possible underlying structures without ambiguity for purposes of economic welfare analysis.

errors in estimation of both. As a result, apparent statistical significance becomes misleading. Arbitrary and seemingly innocuous modeling choices in such problems can explain widely non-convergent research and highly conflicting and misleading policy analysis.

Full structural estimation is possible for such problems by supplementing behavioral data, but imposes demanding data requirements that are likely impractical for most policy analyses, and raise other concerns about bias and dependence on modeling specifications. We propose a more practical approach of modeling a single "value" function in risk problems that combines preferences and perceptions into a generally applicable mean-variance (or, more accurately, mean-risk-premium) criterion. In this case, using a more flexible functional form assures more accuracy in estimation, unlike the case of estimating two multiplicative functions with common variables. This application of the Occam's razor principle (i) avoids the need for unverifiable assumptions to identify preferences and perceptions, (ii) admits standard generalizations that explain behavioral anomalies that have led to concerns about the standard expected utility hypothesis, (iii) supports meaningful policy and economic welfare analysis, and (iv) and permits characterization of risk aversion with empirically meaningful measures that parallel the standard implications of Arrow-Pratt measures that ordinarily apply only under narrow and unverifiable assumptions.

We suggest that this modeling approach can enable meaningful communication and debate among empirical researchers. When estimated concepts depend on arbitrary assumptions that vary among studies, research creates a proliferation of models that provide no reliable basis for meaningful interchange. In contrast, focusing on concepts that are truly econometrically identifiable can provide the basis for empirical knowledge advancement.

References

Alpert, M., and H. Raiffa. (1982). "A Progress Report on the Training of Probability Assessors." In D. Kahneman, P. Slovic and A. Tversky (eds.) *Judgment Bias under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.

Anderson, T.G., T. Bollerslev, F.X. Diebold and H. Ebens. (2001). "The Distribution of Realized Stock Return Volatility." *Journal of Financial Economics* 61, 43-76.

Ariely, D., G. Loesenstein, and D. Prelec. (2003). "Coherent Arbitrariness: Stable Demand Curves without Stable Preferences." *Quarterly Journal of Economics* 118, 73-105.

Barberis, N., M. Huang, and T. Santos. (2001). "Prospect Theory and Asset Prices." *Quarterly Journal of Economics* 116, 1-53.

Benartzi, S. and R. Thaler. (1995). "Myopic Loss Aversion and the Equity Premium Puzzle." *Quarterly Journal of Economics* 110:73-92.

Bernheim, B.D., and A. Rangel. (forthcoming). "Beyond Revealed Preference: Choice Theoretic Foundations for Behavioral Welfare Economics," *Quarterly Journal of Economics*.

Buschena, D., and D. Zilberman. (2009). "Generalized Expected Utility, Heteroscedastic Error, and Path Dependence in Risky Choice." *Journal of Risk and Uncertainty* 20, 67-88.

Camerer, C. (1995). "Individual Decision Making." in J.H. Kagel and A.E. Roth (eds.) *The Handbook of Experimental Economics.* Princeton, NJ: Princeton University Press.

Chavas, J.P. and M.T. Holt. (1996) "Economic Behavior under Uncertainty: A Joint Analysis of Risk Preferences and Technology." *Review of Economics and Statistics* 78, 329-335.

Chavas, J.P. and R.D. Pope. (1985) "Price Uncertainty and Competitive Firm Behavior: Testable Hypotheses from Expected Utility Maximization." *Journal of Economics and Business* 37, 223-235.

Chetty, R. (2008). "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods." Working Paper 14399, National Bureau of Economic Research, Cambridge MA.

Chew, S.H. (1983). "A Generalization of the Quasilinear Mean with Applications to the Measurement of Income Inequality and Decision Theory Resolving the Allais Paradox." *Econometrica*. 51, 1065 – 92.

Chew, S.H. (1989) "Axiomatic Utility Theories with the Betweenness Property." *Annals of Operations Research* 19, 273-298.

De Bondt, W.F.M. and R.H. Thaler. (1985). "Does the Stock Market Overreact?" *Journal of Finance* 40, 793 – 805.

Edwards, W. (1955). "The Prediction of Decisions Among Bets." *Journal of Experimental Psychology* 50, 201-214.

Edwards, W. (1961). "Behavioral Decision Theory." *Annual Review of Psychology* 12, 473-498.

Edwards, W. (1982). "Conservatism in Human Information Processing." In D. Kahneman, P. Slovic and A. Tversky (eds.) *Judgment Bias under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.

Foster, W., and R.E. Just. (1989). "Measuring the Welfare Effects of Product Contamination with Consumer Uncertainty." *Journal of Environmental Economics and Management* 17, 266-283.

Hey, J.D., and C. Orme (1994) Investigating Generalizations of Expected Utility Theory Using Experimental Data. *Econometrica* 62:1291-1326.

Hey, J.D., and E. Strazzera. (1989). "Estimation of Indifference Curves in the Marschak-Machina Triangle: A Direct Test of the 'Fanning Out' Hypothesis." *Journal of Behavioral Decision Making* 2, 239-260.

Hogarth, R.M. and H. J. Einhorn. (1992). "Order Effects in Belief Updating: The Belief-Adjustment Model." *Cognitive Psychology* 24, 1-55.

Jullien, B. and B. Salanie. (2000). "Estimating Preferences under Risk: The Case of Racetrack Bettors." *Journal of Political Economy* 108, 503-530.

Just, D.R., and H.H. Peterson. (2003). "Diminishing Marginal Utility of Wealth and Calibration of Risk in Agriculture." *American Journal of Agricultural Economics* 85:1234-1241.

Just, R.E., L. Calvin, and J. Quiggin. (1999). “Adverse Selection in Crop Insurance: Actuarial and Asymmetric Information Incentives.” *American Journal of Agricultural Economics* 81, 834-849.

Just, R.E., and D.R. Just. (forthcoming). “Global Identification and Tractable Specification Possibilities for Risk Preference Estimation.” *Journal of Econometrics*.

Just, R.E., and Q. Weninger. (1999). “Are Crop Yields Normally Distributed?” *American Journal of Agricultural Economics* 81,287-304.

Kahneman, D. and A. Tversky. (1979). “Prospect Theory: an Analysis of Decision under Risk.” *Econometrica* 47, 263-292.

Katoka, S. (1963). “A Stochastic Programming Model.” *Econometrica* 31, 181-196.

Lichtenstein, S., B. Frischhoff, and L.D. Phillips. (1982). “Calibration of Probabilities: The State of the Art to 1980.” in *Judgment Bias under Uncertainty: Heuristics and Biases*. D. Kahneman, P. Slovic and A. Tversky (eds.) New York: Cambridge University Press.

Lichtenstein, S. and P. Slovic. (1971). “Reversals of Preference Between Bids and Choices in Gambling Decisions.” *Journal of Experimental Psychology* 89, 46-55.

MacCrimmon, K.R., and S. Larsson. (1979) Utility Theory: Axioms versus Paradoxes. In *Expected Utility Hypotheses and the Allais Paradox*. M. Allais and O. Hagen (eds.) Dordrecht: Reidel.

Machina, M.J. (1982). “Expected Utility Analysis Without the Independence Axiom.” *Econometrica* 50, 277-323

Machina, M.J. (1989). “Comparative Statics and Non-expected Utility Preferences.” *Journal of Economic Theory* 47, 393-405.

Machina, M.J. (1995). “Non-Expected Utility and the Robustness of the Classical Insurance Paradigm.” *Geneva Papers on Risk and Insurance Theory* 20, 9-50.

Meyer, J. (1987). Two-moment Decision Models and Expected Utility Maximization, *American Economic Review* 77, 421-430.

Mundlak, Y., and I. Hoch. (1965). “Consequences of Alternative Specifications of Cobb-Douglas Production Functions.” *Econometrica* 33, 814-828.

Murphy, A.H., and R.L. Winkler. (1974). “Subjective Probability Forecasting Experiments in Meteorology: Some Preliminary Results.” *Bulletin of the American Meteorological Society.* 55, 1206-1216.

Preston, M.G. and P. Baratta. (1948). “An Experimental Study of the Auction-Value of an Uncertain Outcome.” *American Journal of Psychology* 61, 183-193.

Pope, R.D. (1980). “The Generalized Envelope Theorem and Price Uncertainty.” *International Economic Review* 21, 75-86.

Quiggin, John. (1982). “A Theory of Anticipated Utility.” *Journal of Economic Behavior and Organization* 3, 323 – 43.

Roy, A.D. (1952). “Safety First and the Holding of Assets.” *Econometrica* 20, 431-449.

Rubenstein, A. (1988). “Similarity and Decision-making under Risk: Is there a Utility Theory Resolution to the Allais Paradox?” *Journal of Economic Theory* 46, 145-153.

Starmer, Chris. (2000). “Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk.” *Journal of Economic Literature* 38, 332-82.

Telser, L.G. (1955-6). “Safety First and Hedging.” *Review of Economic Studies* 23,1-16

Tversky, A., and D. Kahneman. (1982). “Judgment Bias under Uncertainty: Heuristics and Biases.” in *Judgment Bias under Uncertainty: Heuristics and Biases*. D. Kahneman, P. Slovic and A. Tversky (eds.) New York: Cambridge University Press.

Tversky, A., and D. Kahneman. (1992). “Advances in Prospect Theory: Cumulative Representation of Uncertainty.” *Journal of Risk and Uncertainty* 5, 297 – 323.

Verhoef, L.C.G, A.F.J. De Haan and W.A.J. Van Daal. (1994). “Risk Attitude in Gambles with Years of Life.” *Medical Decision Making* 14, 194-200.

Whitehead, J.C., and G.C. Blomquist. 2006. “The Use of Contingent Valuation in Benefit-Cost Analysis.” In A. Alberini and J. Kahn, eds. *Handbook on Contingent Valuation.* Cheltenham UK: Edward Elgar Publishing.

Yaari, M.E. (1987). “The Dual Theory of Choice under Risk.” *Econometrica* 55, 95-115.

Zellner, A., J. Kmenta, and J. Dreze. (1966). “Specification and Estimation of the Cobb-Douglas Production Function.” *Econometrica* 34, 784-795.