J. Econ. Theory 223 (2025) 105940

Contents lists available at ScienceDirect

Journal of Economic Theory

journal homepage: www.elsevier.com/locate/jet

Statistical decision functions with judgment

Simone Manganelli [1]

European Central Bank (ECB), Germany

A R T I C L E I N F O

JEL classification:
C1
C12
C13
C18
D81

Keywords:
Statistical decision theory
Hypothesis testing
Confidence intervals
Ambiguity

A B S T R A C T

A decision maker tests whether the gradient of the loss function evaluated at a judgmental decision is zero, for a given level of significance. If the test does not reject, the decision maker selects the judgmental decision. If the test rejects, the decision maker chooses the action whose gradient is at the boundary of the rejection region. The test is admissible and asymptotically most powerful. The level of significance reflects the decision maker's attitude toward uncertainty. The decision rule is applied to a problem of asset allocation.

1. Introduction

The use of judgment is ubiquitous in decision making. Policy institutions, like central banks, routinely use state of the art econometric models to forecast key economic variables. When forecasts differ from the assessment of the decision makers, they are adjusted with 'expert judgment', either by tinkering with the econometric model or by modifying directly the forecast. It is an *ad hoc* procedure, which clashes with the rigorous foundations of statistical decision theory.

The incorporation of judgment in the decision process should be turned on its head: decision makers first express their judgmental decision and then econometricians recommend whether there is statistical evidence to deviate from it. The surprising implication is that the statistical decision incorporating judgment lies at the boundary of a confidence interval.

Optimality is checked by testing whether, for a strictly convex loss function and level of significance, the gradient evaluated at the judgmental decision is statistically equal to zero. Rejection of the null hypothesis implies that marginal moves decrease the loss function. This holds until the action associated with the closest boundary of the confidence interval of the gradient is reached. Abandoning a judgmental decision for a statistical procedure carries the risk of choosing a worse decision. The level of significance puts an upper bound to the probability of wrongly rejecting a decision when it is optimal.

Maximum likelihood decisions are obtained as a special case of this theory. Such decisions ignore judgment altogether, by setting the level of significance equal to one. In this case, the confidence interval degenerates into a single point, so that any decision, except the maximum likelihood one, is rejected. The statistical interpretation is that the probability that the maximum likelihood decision produces a higher loss than the judgmental decision cannot be bounded away from one.

E-mail address: simone.manganelli@ecb.int.

[1] I would like to thank Pierpaolo Battigalli and three anonymous referees for very insightful comments and useful suggestions. All errors remain my own. The views expressed in this paper are my own and do not necessarily reflect those of the ECB.

https://doi.org/10.1016/j.jet.2024.105940
Received 13 April 2022; Received in revised form 4 May 2024; Accepted 27 November 2024

The contribution of this paper lies at the intersection between statistics and decision theory. Statistical decision theory emerged as a discipline in the 1950's with the works of Wald (1950) and Savage (1954). Berger (1985) provides a comprehensive and accessible review. Works within this tradition are Granger and Machina (2006), Patton and Timmermann (2012) and Elliott and Timmermann (2016). Other contributions include Chamberlain (2000) and Geweke and Whiteman (2006), who deal with forecasting using Bayesian statistical decision theory, Manski (2021) and Manski (2013) (and the references therein), who discusses the current practice of econometrics and its relationship with statistical decision theory, and Christensen et al. (2020), Christensen et al. (2023) and Andrews and Mikusheva (2022) who derive optimal decisions under partial identification. Manganelli (2009) introduces the original idea of incorporating judgment at the beginning of the decision process and proposes the heuristic decision rule of moving to the boundary of the confidence interval. The present paper provides a formal theoretical justification of such a procedure. It also shows that the decision testing judgment optimality is admissible and therefore satisfies the basic rationality principle of not being dominated by any other decision rule.

The paper is structured as follows. Section 2 conveys the main insights of the paper by solving a stylized numerical example. Section 3 sets up the decision environment and introduces the concept of judgment. Judgment is defined as a pair formed by a judgmental decision and a level of significance. It is used to develop the hypothesis testing framework. The key results of this section are that the decision testing optimality of judgment is admissible and the decision compatible with judgment is either the judgmental decision itself or at the boundary of the confidence interval of the sample gradient of the loss function. The fundamental concept behind these results is *conditioning*. Links with the James-Stein estimator and the Bayesian decision are also established in the context of a stylized example. Section 4 provides economic intuition. The level of significance, by putting an upper bound to the probability of wrongly rejecting an optimal decision, can be interpreted as the decision maker's attitude toward uncertainty. Section 5 uses an asset allocation problem as an illustrative example. Section 6 concludes.

2. Heuristics

Before the formal exposition of the model, let us consider a numerical example. The example contains the core idea of the paper and gives the intuition behind the decision with judgment.

Suppose that a random variable X is normally distributed, $X \sim N(\theta, 1)$, the true state of nature is $\theta = 2$, the realization $x = 3$ is observed, the action a is to estimate θ, and the loss function is $L(\theta, a) = 0.5(\theta - a)^2$. Suppose also that the decision maker would take the action $\tilde{a} = 0$, referred to as *judgmental decision*, in the absence of empirical evidence. The question addressed in this paper is if and how the decision maker should modify the judgmental decision $\tilde{a}$, given the empirical evidence $x = 3$ and the set up described in this paragraph.

The problem is depicted in Fig. 1. Let us start from the situation in which the parameter $\theta = 2$ is known. The population loss function is the quadratic function denoted as $L(\theta, a)$, which is minimized by the optimal action $a^0 = 2$. The lower part of the figure reports the population gradient $\nabla_a L(\theta, a)$ as a dotted line, which is zero at $a^0 = 2$, a necessary and sufficient condition for optimality, given the convexity of the loss function. In a situation where the state of nature is known, judgment plays no role, as the optimal decision is available.

This paper is concerned with a situation in which the state of nature is unknown, but the decision maker can make inference about it through the observed realization $x = 3$. In this case, the decision maker can ask if and how the judgmental decision should be modified. Given the empirical evidence, the judgmental decision can be modified, if at all, in the direction provided by the data. In the current setup, the maximum likelihood estimator is X, its estimate is $x = 3$, and the decision associated with it is $\hat{a} = 3$.

It is possible to test for the optimality of $\tilde{a}$ by testing whether its gradient is equal to zero, $H_0 : \nabla_a L(\theta, \tilde{a}) = \tilde{a} - \theta = 0$. To avoid conflating the observed realization x with the potential realization of the underlying random variable X, let us consider the auxiliary random variable $Y \sim N(\theta, 1)$, which is assumed to have the same distribution as X and to be independent of it. A test statistic for this test can be constructed by replacing θ with its maximum likelihood estimator, $\hat{\theta}(Y) = Y$:

$$\nabla_a L(\hat{\theta}(Y), \tilde{a}) = \tilde{a} - Y \sim N(0, 1) \tag{1}$$

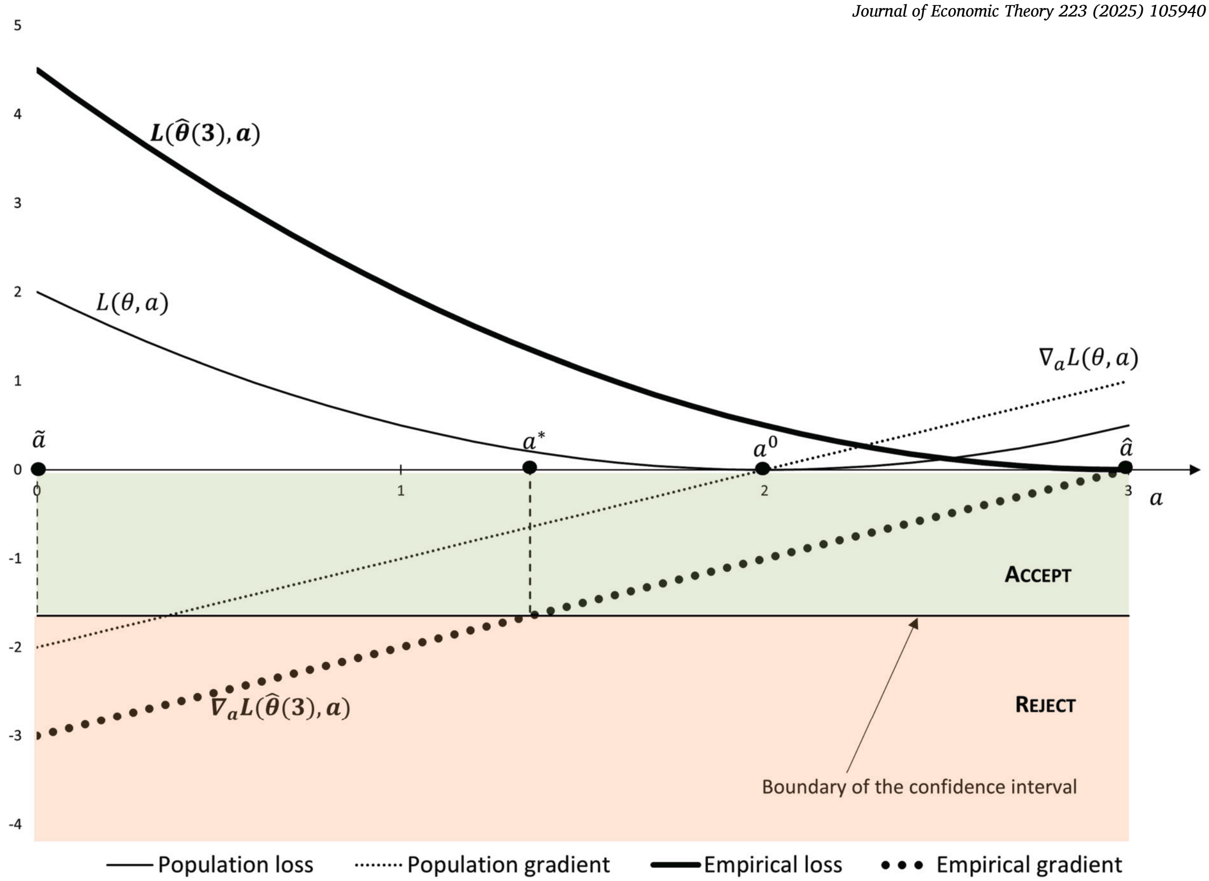
which has a standard normal distribution, under the null hypothesis $H_0 : \tilde{a} - \theta = 0$. Since Y has the same distribution as X, a maximum likelihood estimate can be obtained by replacing the observed sample realization, yielding $\hat{\theta}(3) = 3$.

Implementation of the testing procedure requires the additional choice of a level of significance, α. One can think about the level of significance as the confidence the decision maker has in the judgmental decision $\tilde{a}$: the higher the confidence, the lower α.[2] Alternatively, according to standard statistical jargon, it can be interpreted as the probability of committing Type I errors, that is rejecting the optimality of the decision $\tilde{a}$ when it is optimal.

Choosing $\alpha = 0.10$, the critical value of the test statistic (1) is $c_{\alpha/2} \approx -1.64$, where $c_{\alpha/2}$ is the $\alpha/2$-quantile of the standard normal distribution.[3] This critical value is reported in Fig. 1 as an horizontal line and divides the negative subspace into a rejection and an acceptance region. The thick curve at the top is the empirical loss function, $L(\hat{\theta}(3), a)$, as a function of a. The empirical gradient, $\nabla_a L(\hat{\theta}(3), a)$, is reported at the bottom, as the thick dotted line. The empirical loss is minimized at $\hat{a} = 3$, but recall that the decision maker is interested in minimizing the population loss. Although the population loss cannot be minimized without knowledge of θ, the decision maker can test whether the judgmental decision $\tilde{a} = 0$ minimizes the population loss, by testing whether the empirical

[2] In fact, the probability $1 - \alpha$ is referred to as *confidence level*, so the higher the confidence in the judgmental decision, the higher the confidence level.

[3] The confidence level is divided by 2 because the empirical gradient is known to be either positive or negative, conditional on the observation realization x.



Note: Example of a decision problem with a random variable X normally distributed, $X \sim N(\theta, 1)$, the state of nature $\theta = 2$ and the observed realization $x = 3$. The decision maker needs to choose the action a, for the given loss function $L(\theta, a) = 0.5(\theta - a)^2$. The population loss function $L(\theta, a)$ is minimized by the optimal action $a^0 = 2$. The population gradient, $\nabla_a L(\theta, a)$, is zero at $a^0 = 2$. If the state of nature is unknown, suppose the decision maker would take the judgmental decision $\tilde{a} = 0$, in the absence of empirical evidence. Testing the optimality of $\tilde{a} = 0$ is equivalent to testing whether its gradient is equal to zero. For the given level of significance $\alpha = 0.10$, the critical value of the test statistic $\nabla_a L(\hat{\theta}(Y), \tilde{a}) = \tilde{a} - Y$, where $Y \sim N(\theta, 1)$, is depicted as the horizontal line, which divides the negative subspace into a rejection and an acceptance region. The thick curve at the top is the empirical loss function, $L(\hat{\theta}(3), a)$, evaluated at the maximum likelihood estimate. The empirical gradient, $\nabla_a L(\hat{\theta}(3), a)$, is reported at the bottom as the thick dotted line, and when evaluated at $\tilde{a}$ falls in the rejection region. A negative population gradient means that marginal moves reduce the population loss function. This is true until the boundary of the rejection region is reached, which in the figure is given by the point a^*.

Fig. 1. Example of a Decision with Judgment.

gradient is statistically different from zero. In this example, the empirical gradient evaluated at $\tilde{a}$ falls in the rejection region and therefore the null hypothesis is rejected.

Since under the null hypothesis the population gradient is equal to zero, accepting the alternative implies that the population gradient is negative. A negative population gradient means that marginal moves away from the judgmental decision $\tilde{a}$ in the direction of the maximum likelihood decision $\hat{a}$ reduce the population loss function. This is true until the empirical gradient reaches the boundary of the rejection region, which in the figure is given by the point $a^* \approx 1.36$ and is the only decision compatible with the given judgmental decision $\tilde{a}$ and level of significance α.

Two points deserve further reflection. First, the same test can be carried out for any value of a, by replacing the test statistic (1) with $\nabla_a L(\hat{\theta}(Y), a) = a - Y$ and the null hypothesis with $H_0 : a - \theta = 0$. The actions are not random, conditional on the sample realization $x = 3$, including those between the judgmental decision and the maximum likelihood decision (i.e., $a \in [0, 3]$ in the current example). The direction of update in case of rejection of the judgmental decision is given by the maximum likelihood decision $\hat{a} = 3$. The only element of randomness is in the test statistic $\nabla_a L(\hat{\theta}(Y), a) = a - Y$ and stems from the maximum likelihood estimator $\hat{\theta}(Y)$, not from a. The test statistic is evaluated at the given sample realization and determines whether the null hypothesis is accepted or rejected, which in turn prescribes whether to move away from the current action being tested.

Second, the level of significance α plays a critical role. By changing its value over the interval $[0, 1]$, the resulting decision with judgment spans all the convex combinations between the judgmental decision $\tilde{a}$ and the maximum likelihood decision $\hat{a}$. In the special case $\alpha = 0$ (to be interpreted as the case in which the decision maker knows the state of nature θ), the rejection region in Fig. 1 disappears (as $c_0 = -\infty$) and the null hypothesis that $\tilde{a}$ is optimal is never rejected. In the opposite special case $\alpha = 1$ (to be interpreted as the case in which the decision maker has no prior information about the decision problem), the acceptance region disappears (as $c_{0.50} = 0$), the optimality of every decision is rejected and the decision with judgment is the maximum likelihood decision.

3. Decisions with judgment

This section introduces the concept of judgment and shows how hypothesis testing can be used to arrive at decisions compatible with judgment, first under simplifying assumptions on the data generating process and the loss function, and next under a fairly general asymptotic environment. It also presents a short discussion about the relationship with the James-Stein estimator (Section 3.5) and the Bayesian approach (Section 3.6).

Statistical decision theory has three basic elements: an unknown parameter θ that determines the state of nature, an action a that indicates the decision taken, and a loss function $L(\theta, a)$ that quantifies the loss incurred by the decision maker when the action a is taken and θ is the true state of nature. We argue that judgment forms a fourth necessary element in the construction of the statistical decision problem.

We start by deriving finite sample optimality results of decisions with judgment under the restrictive assumptions that the data are normally distributed and the loss function is quadratic. Next, we show how these optimality results hold in large samples. We denote random variables with upper case letters (X) and their realization with lower case letters (x).

3.1. Judgment

Let us formally introduce each element of the statistical decision problem.

Assumption A1 *(Data Generating Process).* The observed data $x^n \equiv (x_1', \ldots, x_n')'$ are a realization from a distribution P_θ, where $x_t \in \mathbb{R}^v$, $\theta \in \mathbb{R}^p$, and $n, v, p \in \mathbb{N}$. The parameter θ is unknown.

Definition 3.1 *(Action and Loss).* When decision makers choose the **action** $a \in \mathbb{R}^q$, where $q \in \mathbb{N}$, they experience the **loss** $L(\theta, a) : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^+$, a non-negative real number.

An action is optimal if it minimizes the loss function. This concept serves as a benchmark against which the relevant hypothesis is tested and is formally defined as follows.

Definition 3.2 *(Optimality).* For any given θ, an action a^0 is **optimal** if $L(\theta, a^0) \le L(\theta, a)$ for all $a \in \mathbb{R}^q$.

We introduce the following definition of judgment.

Definition 3.3 *(Judgment).* **Judgment** is the pair $A \equiv \{\tilde{a}, \alpha\}$, where $\tilde{a} \in \mathbb{R}^q$ is the **judgmental decision** and $\alpha \in [0, 1]$ is the **level of significance**, or equivalently $1 - \alpha \in [0, 1]$ is the **confidence level**.

Judgment is routinely used in hypothesis testing, for instance when testing whether a regression coefficient is statistically different from zero (with zero in this case playing the role of the judgmental decision), for a given level of significance (usually 1%, 5% or 10%). We say nothing about how the judgmental decision is formed. This question is explored by Tversky and Kahneman (1974) and subsequent research, which however does not address the key question of this paper of how to update the judgmental decision in the light of the sample realization. The choice of the level of significance is discussed in Section 4. For the purpose of this paper, judgment is a primitive to the decision problem, like the loss function.

3.2. Hypothesis testing

For a given judgment, define the following two key actions.

Definition 3.4 *(Maximum Likelihood and Shrinkage Actions).* Denote the maximum likelihood estimator with $\hat{\theta}(X^n)$. The **maximum likelihood action** associated with the sample realization x^n is

$$\hat{a}(x^n) \equiv \arg\min_a L(\hat{\theta}(x^n), a). \tag{2}$$

The **shrinkage action**, shrinking from the judgmental decision $\tilde{a}$ to the maximum likelihood action, is

$$a(\lambda; x^n) \equiv \lambda \hat{a}(x^n) + (1 - \lambda)\tilde{a}, \qquad \lambda \in [0, 1]. \tag{3}$$

Other terms used by econometricians to refer to the maximum likelihood action are *plug-in*, *two-step*, or *as-if* (see Manski (2021) for a critical review of this procedure in the context of statistical decision theory).

Note that $a(1; x^n) = \hat{a}(x^n)$ and $a(0; x^n) = \tilde{a}$. The action $a(\lambda; x^n)$, for any given $\lambda \in [0, 1]$, is observed and therefore non-random at the time the decision is taken, conditional on the sample realization.

The shrinkage action (3) transforms the problem from a q-variate choice of $a \in \mathbb{R}^q$ into a univariate choice of $\lambda \in [0,1]$. This transformation serves mainly a practical purpose. In multivariate problems, the level of significance defines a confidence region, instead of an interval. It would be unclear which action to select from the boundary of the confidence region, as prescribed by the theory developed below. In principle, one could define a q-variate level of significance $\boldsymbol{\alpha} \in [0,1]^q$, so that each element of the action has its own confidence interval, conditional on the other elements. The cost would be an additional burden on the decision maker in the choice of judgment.

The following assumption on the loss function is imposed to make the problem analytically tractable.

Assumption B1 *(Loss Function).* The loss function $L(\theta, a)$ is strictly convex in a and twice continuously differentiable in θ and a.

Under Assumption B1, the *maximum likelihood action* exists and is unique. The decision maker can test whether $a(\lambda; x^n)$ is optimal by testing if the gradient $\nabla_\lambda L(\theta, a(\lambda; x^n))$ is equal to zero. To simplify notation, define the gradient:

$$Z(\theta, \lambda; x^n) \equiv \nabla_\lambda L(\theta, a(\lambda; x^n)), \tag{4}$$

where the conditioning on the sample realization x^n is kept explicit. A test statistic for the gradient can be obtained by replacing θ with its maximum likelihood estimator $\hat{\theta}(X^n)$. The hypothesis to be tested is whether one should move in the direction of $a(1; x^n)$. Since $Z(\hat{\theta}(x^n), 1; x^n) = 0$ and $Z(\hat{\theta}(x^n), \lambda; x^n) < 0$ for $\lambda < 1$ by Assumption B1, one can conclude that higher values of λ decrease the empirical loss function. The decision maker is interested, however, in the population value of the loss function, whose gradient is $Z(\theta, \lambda; x^n)$. If the population gradient is zero, higher values of λ would increase the loss function, rather than decrease it. The null hypothesis to be tested is therefore:

$$H_0 : Z(\theta, \lambda; x^n) \geq 0 \quad \text{vs} \quad H_1 : Z(\theta, \lambda; x^n) < 0. \tag{5}$$

To avoid conflating the potential realizations of X^n with the observed realizations x^n, consider the auxiliary random variable Y^n, which is assumed to have the same distribution as X^n and to be independent of it.

Assumption C *(Auxiliary Random Variable).* The random vector Y^n has the same distribution P_θ of X^n in Assumption A1. The random vectors X^n and Y^n are independent of each other.

The random gradient $Z(\hat{\theta}(Y^n), \lambda; x^n)$ depends not only on the random vector Y^n, but also on the observed sample realization x^n. Under the null hypothesis that λ is optimal, $H_0 : Z(\theta, \lambda; x^n) = 0$, the *p-value* associated with the sample realization $Y^n = x^n$ is $\alpha_\lambda \equiv P[Z(\hat{\theta}(Y^n), \lambda; x^n) \leq Z(\hat{\theta}(x^n), \lambda; x^n)]$. The interpretation is the following. Repeating the hypothetical experiment of drawing independent nv-vectors, y_h^n for $h = 1, \ldots, H$, from the population distribution, the argument of the probability would be true α_λ of the times:

$$\lim_{H \to \infty} H^{-1} \sum_{h=1}^{H} I[Z(\hat{\theta}(y_h^n), \lambda; x^n) \leq Z(\hat{\theta}(x^n), \lambda; x^n)] = \alpha_\lambda. \tag{6}$$

When performing the thought experiment, x^n is held fixed and does not change with y_h^n. By *conditioning* on the data, the potential realizations of the random variable Y^n are not confused with the observed realization x^n.

3.3. Decision

To make further progress, we first assume that the random vectors are normally distributed and the loss function is quadratic. We show that in this case the statistical decision rule testing optimality of judgment is admissible. Next, in subsection 3.4, we use asymptotic arguments to show that the decision is asymptotically optimal.

Assumption A2 *(Normal Data Generating Process).* The distribution P_θ of Assumptions A1 and C is the i.i.d. normal distribution, such that $X_t, Y_t \sim N(\theta, I_v)$, for $t = 1, \ldots, n$, where $\theta \in \mathbb{R}^v$ and I_v is the identity matrix of dimension v.

By Assumption A2, the maximum likelihood estimator $\hat{\theta}(Y^n) \equiv n^{-1} \sum_{t=1}^{n} Y_t$ is normally distributed:

$$\sqrt{n}(\hat{\theta}(Y^n) - \theta) \sim N(0, I_v). \tag{7}$$

Assumption B2 *(Quadratic Loss).* The loss is quadratic with respect to the action: $L(\theta, a) = -a'\theta + \frac{1}{2}a'a$.

By Assumption B2 and Definition 3.4, the gradient becomes:

$$Z(\theta, \lambda; x^n) = (a(1; x^n) - a(0; x^n))'(-\theta + a(\lambda; x^n)). \tag{8}$$

The distribution of the test statistic is therefore:

$$\sqrt{n}\sigma^{-1}\left(Z(\hat{\theta}(Y^n),\lambda;x^n)-Z(\theta,\lambda;x^n)\right)\sim N(0,1), \tag{9}$$

where $\sigma^2 \equiv (a(1;x^n)-a(0;x^n))'(a(1;x^n)-a(0;x^n))$.

Now that a test statistic is available, it is possible to derive a decision rule to test whether an action $a(\lambda;x^n)$ is optimal for any given λ. Let $0<\gamma<1$ and $\Phi(c_\alpha)=\alpha$, where Φ is the cumulative distribution function (cdf) of the standard normal distribution. Define the test function $\psi^A(y^n,\lambda;x^n)$:

$$\psi^A(y^n,\lambda;x^n)=\begin{cases}0 & \text{if } \sqrt{n}\sigma^{-1}Z(\hat{\theta}(y^n),\lambda;x^n)>c_{\alpha/2}\\ \gamma & \text{if } \sqrt{n}\sigma^{-1}Z(\hat{\theta}(y^n),\lambda;x^n)=c_{\alpha/2}\\ 1 & \text{if } \sqrt{n}\sigma^{-1}Z(\hat{\theta}(y^n),\lambda;x^n)<c_{\alpha/2}\end{cases}. \tag{10}$$

Notice that since the sample gradient is negative by construction, the critical region is defined by $\alpha/2$, instead of α.[4]

In an hypothesis testing decision problem, only two actions are possible: the null hypothesis is either accepted or rejected. The connection with the decision problem is the following. The loss function under the null is $L(\theta,a(\lambda;x^n))$, as the action *Accept* in the hypothesis testing decision problem corresponds to the action $a(\lambda;x^n)$ in the original decision problem. The loss function under the alternative is $L(\theta;a(\lambda+\varepsilon;x^n))$ for an infinitesimally small $\varepsilon>0$, as the action *Reject* prescribes to marginally move towards the maximum likelihood estimator and corresponds to the action $a(\lambda+\varepsilon;x^n)$. Because of this correspondence, it is possible to write the decision rule as the test (10), which however should be interpreted as the probability of taking action $a(\lambda+\varepsilon;x^n)$ after observing x^n (see, for instance, Berger (1985) bottom of page 529). The next theorem shows that this decision rule cannot be improved.[5]

Let us first report some additional definitions for convenience.

Definition 3.5 *(Risk Functions).* Define $\bar{L}(\theta,a)\equiv\varepsilon^{-1}L(\theta,a)$ for $\varepsilon>0$ and consider testing the null hypothesis (5) that λ is optimal. The difference between the **risk functions** associated with two decision rules, $\psi_1(Y^n,\lambda;x^n)$ and $\psi_2(Y^n,\lambda;x^n)$, is defined as:

$$\begin{aligned}R(\theta,\psi_1;x^n)-R(\theta,\psi_2;x^n)\equiv&\\ &\equiv\lim_{\varepsilon\to 0}[\bar{L}(\theta,a(\lambda+\varepsilon;x^n))-\bar{L}(\theta,a(\lambda;x^n))]E_\theta[\psi_1(Y^n,\lambda;x^n)-\psi_2(Y^n,\lambda;x^n)]\\ &=Z(\theta,\lambda;x^n)E_\theta[\psi_1(Y^n,\lambda;x^n)-\psi_2(Y^n,\lambda;x^n)].\end{aligned} \tag{11}$$

Definition 3.5 exploits the fact that the loss function is invariant to positive linear transformations. As $\varepsilon\to 0$, the risk of the individual loss function becomes unbounded, because $\bar{L}(\theta,a)$ divides the original loss function by ε. However, the difference between risk functions, which is the relevant concept to define admissibility, is bounded and is a function of the gradient $Z(\theta,\lambda;x^n)$.

Definition 3.6 *(Admissibility).* A decision rule ψ_1 is **inadmissible** if there is another decision rule ψ_2 such that $R(\theta,\psi_2;x^n)-R(\theta,\psi_1;x^n)\le 0$ for all $\theta\in\mathbb{R}^p$ with strict inequality for some θ. The decision ψ_1 is called **admissible** if no such rule ψ_2 exists.

Definition 3.7 *(Completeness).* A class $\mathbb{C}$ of decision rules for testing the optimality of the null hypothesis (5) is **essentially complete** if, for any decision rule $\psi_2\notin\mathbb{C}$, there is a decision rule $\psi_1\in\mathbb{C}$ such that $R(\theta,\psi_2;x^n)-R(\theta,\psi_1;x^n)\ge 0$ for all $\theta\in\mathbb{R}^p$.

A standard result in statistical decision theory is that if an admissible decision rule ψ is not in an essentially complete class $\mathbb{C}$, then there exists a decision rule $\psi'\in\mathbb{C}$ which is equivalent to ψ (see for instance Lemma 2, page 522, Berger (1985)). An essentially complete class does not necessarily contain all admissible decisions, but it contains all admissible risk functions.

Theorem 3.1 *(Complete Class). Given Assumptions A2, B2 and C, the class of tests in* (10) *forms an essentially complete class and any test of the form* (10) *is admissible.*

Proof. See Appendix. □

The admissibility result is obtained by applying Karlin-Rubin theorem to the test function (10). It follows from two facts. First, even though the action $a(\lambda;x^n)$ may be a vector, the tested hypothesis is about a scalar, the derivative of the loss function with respect to λ. Second, the randomness of the decision rule stems from the test function $\psi^A(Y^n,\lambda;x^n)$, while the corresponding actions in case of rejection or non rejection are not random, because they condition on the observed realization x^n.

The next theorem finally derives the decision compatible with judgment.

[4] It is known *ex ante* that the random variable $Z(\hat{\theta}(Y^n);\lambda;x^n)$ evaluated at $Y^n=x^n$ is negative. This information must be incorporated into the computation of the probability. The probability statement incorporating this information is $\Pr[Z(\hat{\theta}(Y^n);\lambda;x^n)<Z(\hat{\theta}(x^n);\lambda;x^n)|Z(\hat{\theta}(Y^n);\lambda;x^n)<0]$, which leads to a critical region defined by $\alpha/2$.

[5] Recall that, given the convexity of the loss function, testing the null hypothesis that a given action $a(\lambda;x^n)$ is optimal is equivalent to testing whether the population gradient evaluated at that action is equal to zero. In case of rejection, the alternative says that the population gradient evaluated at that action is different from zero. Therefore, under the alternative hypothesis, a marginal move away from the tested action, $a(\lambda+\varepsilon;x^n)$ for a sufficiently small $\varepsilon>0$, decreases the population loss function.

Theorem 3.2 *(Decision compatible with judgment). Consider the class of decision rules $\delta(x^n) : \mathbb{R}^{vn} \to [0,1]$. Given Assumptions A2, B2 and C, the decision compatible with the judgment $A = \{\tilde{a}, \alpha\}$ is $\delta^A(x^n) = \hat{\lambda}$, resulting in the action $a(\hat{\lambda}; x^n)$ from (3), where $\hat{\lambda} = \max\{0, \lambda^*\}$ and λ^* solves $\psi^A(x^n, \lambda^*; x^n) = \gamma$ in (10) if $\tilde{a} \neq \hat{a}(x^n)$, and $\hat{\lambda} \in [0,1]$ if $\tilde{a} = \hat{a}(x^n)$.*

Proof. See Appendix. □

The decision of Theorem 3.2 is the only decision compatible with the given judgment and the given realization of the test statistic. The compatibility with the judgment is assessed with the test of Theorem 3.1, which contains the only random element of the statistical decision problem, conditional on the sample realization x^n: the maximum likelihood estimator $\hat{\theta}(Y^n)$. The test function (10) is testing the null hypothesis that $a(\hat{\lambda}; x^n)$ is optimal and is evaluated at x^n: since Y and X have the same distribution, the sample realization x^n can be used to construct the maximum likelihood estimate $\hat{\theta}(x^n)$. The action $a(\hat{\lambda}; x^n)$ is not random, conditional on the sample realization x^n, and therefore it is meaningless to talk about admissibility with respect to this action. The correct interpretation is that there is a probability not greater than α of incorrectly rejecting $a(\hat{\lambda}; x^n)$ as optimal decision, following the logic described in equation (6). Theorem 3.1 shows that the test (10) used to arrive at this decision is admissible.

3.4. Asymptotic optimality

Optimality results in finite sample situations are usually restricted to special cases, such as those illustrated by Assumptions A2 and B2. More general results can be obtained by allowing the sample size to increase and applying optimality to the limiting case. It turns out that the limiting situation is similar to that described by Theorem 3.1. One feature of these results is that the null hypothesis $H_0 : Z(\theta_0, \lambda; x^n) \geq 0$ is tested against a sequence of alternatives $H_1 : Z(\theta_0 + hn^{1/2}, \lambda; x^n) < 0$ that tend to H_0 as the sample size grows, as otherwise the power of the test would converge asymptotically to 1. We refer to Lehmann and Romano (2005), Chapters 11-13, for a detailed discussion which builds on Le Cam (1986). Similar approaches are taken by Christensen et al. (2020), Christensen et al. (2023) and Andrews and Mikusheva (2022) to derive optimal decisions in non-standard settings.

Before stating the theorem, we need an additional assumption about the asymptotic behavior of the maximum likelihood estimator.

Assumption D *(Consistency and asymptotic normality).* The maximum likelihood estimator is consistent and asymptotically normally distributed:

$$n^{1/2}(\hat{\theta}(Y^n) - \theta) \xrightarrow{d} N(0, \Sigma). \tag{12}$$

See, for instance, Newey and McFadden (1994) for the set of conditions needed for Assumption D to hold.

Let us define the quantities $a^0(\theta) \equiv \arg\min_a L(\theta, a)$ and $a(\lambda; \theta) \equiv \lambda a^0(\theta) + (1-\lambda)\tilde{a}$ for $\lambda \in [0,1]$, which are the optimal and shrinkage actions in population, respectively. These quantities are needed to define the population gradient, a key ingredient of the asymptotic result:

$$Z(\theta, \lambda) \equiv \nabla_\lambda L(\theta, a(\lambda; \theta)). \tag{13}$$

We can now state the result about asymptotic optimality of the test (10).

Theorem 3.3 *(Asymptotic Optimality). Suppose the Assumptions A1, B1, C and D hold. Consider testing the null hypothesis $H_0 : Z(\theta, \lambda_n; x^n) \geq 0$ against the alternative $H_1 : Z(\theta, \lambda_n; x^n) < 0$, for a given judgment $A = \{\tilde{a}, \alpha\}$. Let θ_0 be such that $Z(\theta_0, \lambda_n; x^n) = 0$ and $\theta = \theta_0 + hn^{-1/2}$. Then the class of test ψ^A in (10) with $\sigma^2 \equiv \nabla'_\theta Z(\theta_0, 1)\Sigma\nabla_\theta Z(\theta_0, 1)$ is pointwise asymptotically level α and for any h such that $h'\nabla_\theta Z(\theta_0, \lambda_n; x^n) < 0$ its limiting power is*

$$\lim_n E_{\theta_0 + hn^{-1/2}}[\psi^A(Y^n, \lambda_n; x^n)] = \Phi(c_\alpha - \sigma^{-1} h' \nabla_\theta Z(\theta_0, 1)), \tag{14}$$

where $Z(\theta_0, 1)$ is defined in (13).

Proof. See Appendix. □

Given the dependence of the gradient on the sample realization x^n, the notation λ_n is used to ensure that the null hypothesis $Z(\theta_0, \lambda_n; x^n) = 0$ holds for all n. Notice that the condition $Z(\theta_0, \lambda_n; x^n) = 0$ is equivalent to $a^0(\theta_0) = a(\lambda_n; x^n)$.

For fixed n, the test based on the (unknown) exact distribution is replaced by an approximating test based on the normal distribution. The decision compatible with judgment of Theorem 3.2 extends directly to the test based on this approximation and is determined by the corresponding $\hat{\lambda}_n$.

3.5. Stein's paradox

Suppose that $X \sim N(\theta, I_p)$, with dimension p, only one realization $x \neq 0$ is observed (that is, $n = 1$), the action a is to estimate θ, and the loss function is $L(\theta, a) = \sum_{i=1}^p (\theta_i - a_i)^2/2$. Suppose also that the judgmental decision is $\tilde{a} = 0$. In this case, $\hat{\theta}(x) = x$,

$a(1;x)=x$, $a(\lambda;x)=\lambda x$ and $a(0;x)=0$. The population gradient, conditional on the observation x, is $Z(\theta,\lambda;x)=(\lambda x-\theta)'x$, while the test statistic is $Z(\hat{\theta}(Y),\lambda;x)=(\lambda x-Y)'x$, where $Y\sim N(\theta,I_p)$ has the same distribution as X. The distribution of the test statistic is $Z(\hat{\theta}(Y),\lambda;x)\sim N(\lambda x'x-\theta'x,x'x)$. Under the null hypothesis $H_0:\theta=\lambda x$, the critical value of the test statistic is $c_{\alpha/2}\sqrt{x'x}$. Then, according to Theorem 3.2, the decision with judgment is given by $\delta^A(x)=\hat{\lambda}x$, where $\hat{\lambda}=\max\{0,\lambda^*\}$ and $\lambda^*=1+c_{\alpha/2}/\sqrt{x'x}$.

When p, the dimension of θ, is greater than 2, it is well known that the maximum likelihood estimator $\hat{\theta}(X)=X$ is inadmissible (Stein (1956)). When $\alpha=1$ in the above example, $\hat{\lambda}=1$ (any null hypothesis is rejected, since in this case $c_{\alpha/2}=0$ and the confidence interval degenerates into the singleton 0) and the decision with judgment $a(\hat{\lambda};x)=x$ corresponds to the maximum likelihood decision. By Theorem 3.2, it is the only decision compatible with the judgment $A=\{0,1\}$.

The James-Stein estimator, on the other hand, is obtained by setting $\lambda=1-(p-2)/x'x$. This corresponds to the decision with judgment when $c_{\alpha/2}=-(p-2)/\sqrt{x'x}$, provided that $-(p-2)/x'x$ is greater than -1.[6] This in turn corresponds to the level of significance $\alpha=2\Phi^{-1}(-(p-2)/\sqrt{x'x})<1$. The James-Stein decision is therefore not compatible with the given judgment of the current example, $A=\{0,1\}$.

3.6. Connection with Bayesian decisions

The Bayesian approach assumes that the decision maker uses subjective information in the form of a prior distribution over the unknown parameter θ. Let us further simplify the example of the previous subsection, so that we can analytically derive the connection between the two decisions. Assume that $X\sim N(\theta,1)$ is univariate, so that $\theta\in\mathbb{R}$ and $L(\theta,a)=(\theta-a)^2/2$. Let us also assume that the observed realization is $x\neq 0$.

In this case, the Bayesian decision with prior distribution $N(0,1)$ over θ is $a(\delta^B(x);x)=x/2$, where the superscript B stands for Bayesian. The decision rule incorporating judgment $A=\{0,\alpha\}$, on the other hand, is $a(\delta^A(x);x)=\hat{\lambda}x$, where $\hat{\lambda}=\max\{0,\lambda^*\}$ with $\lambda^*=1+c_{\alpha/2}/|x|$.

The link between Bayesian decisions and decisions with judgment is given by the link between priors, posteriors and judgment $A\equiv\{\tilde{a},\alpha\}$ and is obtained by imposing that $a(\delta^B(x);x)=a(\delta^A(x);x)$. In the current example, for the Bayesian decision associated with a standard normal prior there is the following corresponding choice of $\tilde{a}$ and α, which produces an observationally equivalent decision with judgment:

$$\tilde{a}=0, \tag{15}$$

$$\alpha(x)=2\Phi\left[\Phi^{-1}[\tilde{\alpha}(x)/2]/2\right], \tag{16}$$

where $\Phi(\cdot)$, as usual, denotes the cdf of the standard normal distribution and $\tilde{\alpha}(x)\equiv 2\Phi(-|x|)$ is the *p-value* associated with the sample realization x, under the null hypothesis that $\tilde{a}=0$ is optimal.[7]

This equivalence results reveals that Bayesian decisions can be interpreted as decisions with judgment, when judgment is appropriately defined. According to this interpretation, Bayesian decisions first test the null hypothesis that the decision associated with the prior (i.e. the judgmental decision $\tilde{a}$ in (15)) is optimal and next, in case of rejection, select the boundary of the confidence interval associated with the confidence level $1-\alpha(x)$, following the reasoning of Theorem 3.2.

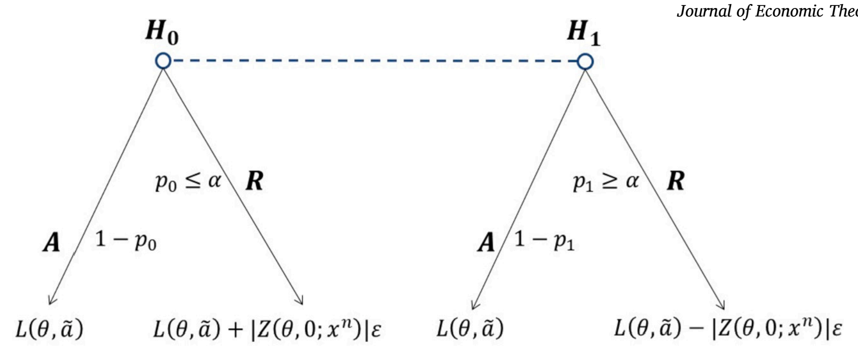
The level of significance depends on x in (16), highlighting that Bayesian decisions condition on the data not only in the hypothesis testing as discussed in the paragraph containing equation (6), but also in the choice of α. It is easy to show that for this choice of the level of significance, the *p-value* is always lower than the level of significance: $\tilde{\alpha}(x)<\alpha(x)$ for all $x\neq 0$. The implication is that the Bayesian decision always rejects the judgmental decision $\tilde{a}$ and therefore the probability that it performs worse than $\tilde{a}=0$, when it is optimal, cannot be bounded away from 1. In other words, Bayesian decisions have no control on Type I errors, when testing if the judgmental decision is optimal. This is linked to the fact that Bayesian decision makers exhibit a neutral attitude towards uncertainty.[8] Suitable choices of the level of significance can cover cases in which the decision maker is averse to uncertainty, as in robust Bayesian decisions where decision makers minimize the loss function using the worst possible prior from a given class of priors (see Gilboa and Marinacci (2013) for a review of the literature). We refer to Manganelli (2023) for a more comprehensive discussion about the relationship between decisions with judgment and Bayesian decisions.

---

[6] If this condition does not hold, the James-Stein shrinkage factor would be negative, which defies economic intuition, as the decision would shrink away from the maximum likelihood decision. The James–Stein estimator dominates the mean even when shrinkage factor is negative, when conditioning on the sample realization is not taken into account. There is also the positive-part James–Stein estimator which truncates negative shrinkage to zero.

[7] Going in the opposite direction, from the decision with judgment to the Bayesian decision, requires to search through all the possible prior distributions such that (1) the Bayesian decision based on the prior is equal to the judgmental decision and (2) the Bayesian decision based on the posterior is equal to the decision with judgment, i.e. $a(\delta^B(x);x)=a(\delta^A(x);x)$. The existence and uniqueness of such prior and posterior is beyond the scope of the current discussion.

[8] An equivalent term used in the literature is ambiguity. Ellsberg (1961) has shown that people prefer bets where the odds of an event are known (*risky events*) to bets where the odds are unknown (*uncertain or ambiguous events*). This type of behavior is incompatible with the Bayesian approach, which relies on a single prior distribution over the state space.



Note: A decision maker with judgmental decision $\tilde{a}$, level of significance α and loss function $L(\theta, \tilde{a})$ faces the depicted statistical decision problem. For a given estimator $\hat{\theta}(Y^n)$ of θ, the rule tests whether marginal ($\varepsilon > 0$) deviations from $\tilde{a}$ are warranted, by accepting (A) or rejecting (R) the null hypothesis that the gradient of the loss function is non-negative, $H_0 : Z(\theta, 0; x^n) \ge 0$. It will not decrease the loss if the hull hypothesis is true (node H_0) and it will decrease the loss if the hull hypothesis is false (node H_1). The dashed line connecting H_0 and H_1 represents uncertainty, as the decision maker cannot distinguish between the two parts of the tree and no probability can be attached to them. By choosing α, she can control the probability p_0 of increasing the loss function, in case H_0 is true. The level of significance α provides also the lower bound to the probability p_1 of correctly deviating from $\tilde{a}$ in case H_1 is true.

Fig. 2. Statistical Decision Tree.

4. Economic intuition

To understand the intuition behind Theorem 3.1, consider that the null hypothesis (5) for $\lambda = 0$ is a statement about the population gradient evaluated at the judgmental decision $\tilde{a}$. It says that *marginal* moves from $\tilde{a}$ in the direction of $\hat{a}(x^n)$ do not decrease the loss function. If it is not rejected at the given level of significance α, the decision maker chooses $\tilde{a}$. Rejection of the null hypothesis, on the other hand, implies accepting the alternative, which states that *marginal* moves away from $\tilde{a}$ decrease the loss function.

The decision problem is depicted in Fig. 2. The decision maker has no information to distinguish the left part of the decision tree, denoted by the node H_0, from the right part, denoted by H_1. Under H_0, the null hypothesis (5) is true, so that any deviation from the judgmental decision $\tilde{a}$ results in the higher loss $L(\theta, \tilde{a}) + |Z(\theta, 0; x^n)|\varepsilon$ for sufficiently small $\varepsilon > 0$. Under H_1, the null hypothesis is false, and a marginal ε move away from $\tilde{a}$ results in the lower loss $L(\theta, \tilde{a}) - |Z(\theta, 0; x^n)|\varepsilon$. The dash line connecting the two nodes represents true uncertainty for the decision maker, in the sense that it is not possible to attach any probability to being in H_0 or in H_1. The decision maker can choose the level of significance α, which puts an upper bound to the probability that the null is wrongly rejected when it is true, and a lower bound to the probability of correctly rejecting H_0 when it is false.

In case of rejection, the preferred decision $a(\hat{\lambda}; x^n)$ is the action which lies at the boundary of the $(1 - \alpha)$ confidence interval of the gradient. Other actions would not be compatible with the level of significance α of the decision maker. In fact, actions closer to the original judgmental decision $\tilde{a}$ are rejected at the level of significance α, while actions further away are wrongly rejected with a probability greater than α.

The level of significance α determines the willingness of the decision maker to engage in the statistical bet. A decision maker who is uncertainty neutral chooses $\alpha = 1$. When $\alpha = 1$ the confidence interval degenerates into a single point and the null hypothesis that $\tilde{a}$ is optimal is always rejected (that is, there is no control on Type I errors). The decision consistent with this level of significance is the *maximum likelihood* decision $a(1; x^n)$. At the opposite end of the spectrum, a decision maker with an extreme aversion to uncertainty chooses $\alpha = 0$. When $\alpha = 0$ the confidence interval degenerates into the entire real line and the null hypothesis that $\tilde{a}$ is optimal is never rejected. This corresponds to the minmax decision $a(0; x^n)$ relative to the judgment $A = \{\tilde{a}, \alpha\}$, in the sense that it guarantees a loss not greater than $L(\theta, \tilde{a})$. The decision avoids any statistical risk (that is, no Type I errors are committed). An intermediate case is represented by the *subjective classical* estimator of Manganelli (2009), which sets $\alpha \in (0, 1)$ and gives the decision $a(\hat{\lambda}; x^n)$ of Theorem 3.2.

There is a trade-off associated with the choice of the level of significance, which is the one associated with Type I and Type II errors in hypothesis testing. Lower values of α imply a lower probability of wrongly rejecting the null hypothesis, but also a lower probability of correctly rejecting it. It is up to the decision maker to solve this trade-off. The choice of the level of significance depends on the decision problem at hand and the confidence that decision makers have on their own judgmental decision. Notice that it is impossible not to choose. Any decision maker facing a statistical decision problem is forced to choose a level of significance.

5. An asset allocation decision problem

This section implements the decision with judgment, solving a portfolio allocation problem. The empirical implementation of the mean-variance asset allocation model introduced by Markowitz (1952) has puzzled economists for a long time. Despite its theoretical success, standard estimators of the portfolio weights produce volatile asset allocations with poor out-of-sample performance (Brandt (2009)). This paper takes a different perspective on this problem, by starting with a judgmental portfolio allocation and testing whether its performance can be improved.

To implement the statistical decision rule of Theorem 3.2, we take a monthly series of closing prices for the EuroStoxx50 index, from February 1987 until September 2019. EuroStoxx50 covers the 50 leading Blue-chip stocks for the Eurozone. The data is taken

Table 1
Asset allocation decisions.

		α			
		0	0.01	0.10	1
	0	0	0	0	0.19
$\tilde{a}$	0.5	0.5	0.5	0.4	0.19
	1	1	0.53	0.4	0.19

Note: Share of wealth invested in the monthly Eurostoxx50 index, according to alternative choices of judgmental decision ($\tilde{a}$) and level of significance (α). The case in which $\alpha = 1$ always ignores any judgmental decision and chooses the decision associated with the maximum likelihood estimate. The case in which $\alpha = 0$ always ignores any statistical evidence and chooses the judgmental decision. The case in which $\alpha \in (0, 1)$ results in decisions which shrink toward the maximum likelihood decision, provided there is sufficient statistical evidence to move away from the judgmental decision.

from DataStream. The closing prices are converted into period log returns, $x^n \equiv (x_1, \ldots, x_n)'$, for a total of $n = 392$ monthly observations. Assume for simplicity $E_t(X_{t+1}) = \theta_1$ and $V_t(X_{t+1}) = \theta_2$, that is both first and second moments are not time varying, and define $\theta \equiv [\theta_1, \theta_2]'$. The methodology can be readily applied to cases where the conditional mean and variance are time varying.

Consider an investor with a quadratic utility function $U(W) = W - 0.5bW^2$ with $b > 0$ and $W < 1/b$. The decision is about the fraction a of cash w_n available at the end period n to be invested in the stock market. Assuming that the return on cash is zero, the monthly portfolio returns are ax_i, for $i = 1, \ldots, n$, so that the random variable at the end of period n is $W = w_n a X_{n+1}$. The loss function is the negative of the expected utility and is, up to a positive linear transformation, $L(\theta, a) = -a\theta_1 + 0.5bw_n a^2(\theta_2 + \theta_1^2)$. Assuming a normal likelihood, the decision associated with the maximum likelihood estimate is $\hat{a}(x^n) = \hat{\theta}_1(x^n)/[(bw_n(\hat{\theta}_2(x^n) + \hat{\theta}_1(x^n)^2)]$, where $\hat{\theta}_1(x^n) = n^{-1}\sum_{i=1}^n x_i$ and $\hat{\theta}_2(x^n) = n^{-1}\sum_{i=1}^n (x_i - \hat{\theta}_1(x^n))^2$. Further assuming correct specification, the asymptotic variance-covariance matrix is (see Newey and McFadden (1994)):

$$\hat{\Sigma} = \left[n^{-1} \sum_{i=1}^{n} s(x_i, \hat{\theta}(x^n)) s(x_i, \hat{\theta}(x^n))' \right]^{-1},$$

$$s(x_i, \hat{\theta}(x^n)) = \begin{bmatrix} (x_i - \hat{\theta}_1(x^n))\hat{\theta}_2(x^n)^{-1} \\ -0.5\hat{\theta}_2(x^n)^{-1} + 0.5(x_i - \hat{\theta}_1(x^n))^2 \hat{\theta}_2(x^n)^{-2} \end{bmatrix}.$$

The gradient and its derivative are:

$$Z(\hat{\theta}(x^n), \lambda; x^n) = -(\hat{a}(x^n) - \tilde{a})\hat{\theta}_1(x^n) + bw_n a(\lambda; x^n)(\hat{a}(x^n) - \tilde{a})(\hat{\theta}_2(x^n) + \hat{\theta}_1(x^n)^2),$$

$$\nabla_\theta Z(\hat{\theta}(x^n), 1; x^n) = \begin{bmatrix} -(\hat{a}(x^n) - \tilde{a}) + 2bw_n a(1; x^n)(\hat{a}(x^n) - \tilde{a})\hat{\theta}_1(x^n) \\ bw_n a(1; x^n)(\hat{a}(x^n) - \tilde{a}) \end{bmatrix}.$$

There are now all the elements to compute $\hat{\sigma}^2$, as defined in Theorem 3.3.

The decision rule of Theorem 3.2 is implemented by choosing $bw_n = 9$ and the choices of $A = \{\tilde{a}, \alpha\}$ reported in Table 1. By choosing $\alpha = 1$, the decision maker ignores any judgmental decision and selects the decision associated with the maximum likelihood estimate. In the current exercise, this corresponds to investing 19% of the portfolio in the stock index and keeping the rest in cash. At the other extreme, by choosing $\alpha = 0$, the decision maker ignores any statistical evidence and selects the judgmental decision. This can be seen from the fact that the decisions in the column under $\alpha = 0$ are identical to the corresponding $\tilde{a}$.

Intermediate choices of the level of significance, $\alpha \in (0, 1)$, result in decisions which shrink toward the maximum likelihood decision, provided there is sufficient statistical evidence to move away from the judgmental decision. The null hypothesis that $\tilde{a} = 0$ is optimal is not rejected at the 10% level of significance, and therefore the decision coincides with $\tilde{a}$. Notice that this finding explains the lack of participation in the stock market, even though the standard expected utility theory predicts that all agents should always invest some fraction of their wealth in the risky asset. Investors averse to uncertainty prefer not to invest in the stock market if the available statistical evidence is not strong enough. This explanation is consistent with a larger body of literature which explains the lack of participation with the assumption that investors view stock returns as ambiguous (Epstein and Schneider (2010)).

Investors prefer also not to move away from the judgmental decision $\tilde{a} = 0.5$ at the 1% level of significance. This judgmental decision is however rejected at 10%, leading to an investment in the stock market of 40% of the overall portfolio. Notice that $\tilde{a} = 1$ is also rejected and leads to the same decision as the one associated with $\tilde{a} = 0.5$, when $\alpha = 0.10$.

6. Conclusion

Judgment plays an important role not just for individuals, but also in policy institutions. Most policy decisions are shaped by the judgment of policy makers. When advising a policy maker, the econometrician can test whether the preferred judgmental decision is supported by models and data. If not, the decision incorporating judgment is always at the closest boundary of the confidence interval. The probability of wrongly rejecting the judgmental decision is bounded by the given level of significance. The decision rule is admissible and is obtained by properly conditioning on the observed sample realization.

The level of significance reflects the decision maker's attitude toward uncertainty. Decision makers with extreme aversion to uncertainty always follow their own judgmental decision and ignore the advice of the econometrician. At the other extreme, decision makers indifferent to uncertainty ignore their judgment and always choose the maximum likelihood decision. Policy makers engaging in statistical decision making are likely characterized by low, but not extreme, levels of significance.

CRediT authorship contribution statement

Simone Manganelli: Conceptualization.

Declaration of competing interest

None

Appendix A. Proofs

Proof of Theorem 3.1. We verify that the assumptions of Theorem 5 on page 530 of Berger (1985) are satisfied.

By Assumptions A2, B2 and C, it follows that

$$\mathcal{X} \sim N(\vartheta, 1),$$

where $\mathcal{X} \equiv \sqrt{n}\sigma^{-1} Z(\hat{\theta}(Y^n), \lambda; x^n)$ and $\vartheta \equiv \sqrt{n}\sigma^{-1} Z(\theta, \lambda; x^n)$.

Since $\mathcal{X}$ is normally distributed, it has a monotone likelihood ratio and $\{x : f(x|\vartheta) > 0)\}$ is independent of ϑ, where $f(x|\vartheta)$ is the probability density function of $\mathcal{X}$.

Finally, denote with λ_n the choice of λ at time n and let $\theta_0 = a(\lambda_n; x^n)$, so that $Z(\theta_0, \lambda_n; x^n) = 0$, by equation (8). Define $\vartheta_0 = \sqrt{n}\sigma^{-1} Z(\theta_0, \lambda_n; x^n)$. The loss function is as in equation (8.9) of Berger (1985) (with inverted signs, as we are testing $H_0 : \vartheta \geq \vartheta_0$ vs $H_1 : \vartheta < \vartheta_0$), since for $\varepsilon > 0$

$$\lim_{\varepsilon \to 0} \varepsilon^{-1}[L(\theta, a(\lambda_n + \varepsilon; x^n)) - L(\theta, a(\lambda_n; x^n))] < 0 \quad \text{if } \vartheta < \vartheta_0$$
$$\lim_{\varepsilon \to 0} \varepsilon^{-1}[L(\theta, a(\lambda_n + \varepsilon; x^n)) - L(\theta, a(\lambda_n; x^n))] > 0 \quad \text{if } \vartheta > \vartheta_0,$$

because by definition $\lim_{\varepsilon \to 0} \varepsilon^{-1}[L(\theta, a(\lambda_n + \varepsilon; x^n)) - L(\theta, a(\lambda_n; x^n))] = Z(\theta, \lambda_n; x^n)$ and we exploit the property that the loss function is invariant to a positive linear transformation. □

Proof of Theorem 3.2. If $\psi^A(x^n, 0; x^n) = 0$, the null hypothesis $H_0 : Z(\theta_0, 0; x^n) = 0$ is not rejected at the given level of significance α. $\tilde{a}$ is therefore retained as the chosen action.

If $\psi^A(x^n, 0; x^n) = 1$, the null hypothesis is rejected. Let λ^* be the value satisfying $\sqrt{n}\sigma^{-1} Z(\hat{\theta}(x^n), \lambda^*; x^n) = c_{\alpha/2}$. Given Assumptions A2 and B2, this value exists and is unique. Denote with $a(\bar{\lambda}; x^n)$ the chosen action and suppose by contradiction that $\bar{\lambda} \neq \lambda^*$. If $\bar{\lambda} < \lambda^*$, this implies that $\sqrt{n}\sigma^{-1} Z(\hat{\theta}(x^n), \bar{\lambda}; x^n) < c_{\alpha/2}$, that is $H_0 : Z(\theta_0, \bar{\lambda}; x^n) = 0$ should be rejected. This decision is not compatible with the decision maker's judgment, since it does not reject despite having a *p-value* lower than the chosen level of significance α. If $\bar{\lambda} > \lambda^*$, continuity implies that it exists $\varepsilon > 0$ such that the null $H_0 : Z(\theta_0, \bar{\lambda} - \varepsilon; x^n) = 0$ is rejected at the given level of significance α, even though $\sqrt{n}\sigma^{-1} Z(\hat{\theta}(x^n), \bar{\lambda} - \varepsilon; x^n) > c_{\alpha/2}$. This decision is also not compatible with the decision maker's judgment, because it rejects with a probability higher than the chosen level of significance α. The chosen action must therefore be $\bar{\lambda} = \lambda^*$. □

Proof of Theorem 3.3. By Assumption B1, it exists θ_0 such that $Z(\theta_0, \lambda_n; x^n) = 0$. We verify that the conditions of Theorem 13.5.1 of Lehmann and Romano (2005) are satisfied.

The i.i.d. assumption on $Y_1, \dots, Y_n$ and the quadratic mean assumption are replaced by Assumption D. Together with Assumptions A1, B1 and C, a mean value expansion gives:

$$\sqrt{n}[Z(\hat{\theta}(Y^n), \lambda_n; x^n) - Z(\theta_0, \lambda_n; x^n)] = \nabla_\theta Z(\bar{\theta}(Y^n), \lambda_n; x^n)\sqrt{n}(\hat{\theta}(Y^n) - \theta_0)$$
$$\xrightarrow{d} N(0, \sigma^2),$$

where $\bar{\theta}(Y^n)$ lies in between $\hat{\theta}(Y^n)$ and θ_0. Notice that $\nabla_\theta Z(\bar{\theta}(Y^n), \lambda_n; x^n) \xrightarrow{P} \nabla_\theta Z(\theta_0, 1)$, because $H_0 : a^0(\theta_0) = a(\lambda_n; x^n)$ and Assumption D imply $\lim_{n\to\infty} \lambda_n = 1$. All the conditions of Theorem 13.5.1 (ii) are satisfied and the result follows. □

Appendix B. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jet.2024.105940.

Data availability

I have shared data and Matlab codes in the 'Attach file' step.

References

Andrews, I., Mikusheva, A., 2022. Optimal decision rules for weak GMM. Econometrica 90 (2), 715–748.
Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis, 2nd ed. Springer-Verlag, New York.
Brandt, M., 2009. Portfolio choice problems. In: Ait-Sahalia, Y., Hansen, L.P. (Eds.), Handbook of Financial Econometrics. North Holland.
Chamberlain, G., 2000. Econometrics and decision theory. J. Econom. 95 (2), 255–283.
Christensen, T., Moon, H.R., Schorfheide, F., 2020. Robust forecasting. arXiv:2011.03153.
Christensen, T., Moon, H.R., Schorfheide, F., 2023. Optimal decision rules when payoffs are partially identified. arXiv:2204.11748.
Elliott, G., Timmermann, A., 2016. Economic Forecasting. Princeton University Press.
Ellsberg, D., 1961. Risk, ambiguity and the savage axioms. Q. J. Econ. 75 (4), 643–669.
Epstein, L., Schneider, M., 2010. Ambiguity and asset markets. Annu. Rev. Financ. Econ. 2, 315–346.
Geweke, J., Whiteman, C., 2006. Bayesian forecasting. In: Elliott, C.G.G., Timmermann, A. (Eds.), Handbook of Economic Forecasting, vol. 1. Elsevier, pp. 3–80.
Gilboa, I., Marinacci, M., 2013. Ambiguity and the Bayesian paradigm. In: Acemoglu, D., Arellano, M., Dekel, E. (Eds.), Advances in Economics and Econometrics: Theory and Applications. Cambridge University Press, New York.
Granger, C., Machina, M., 2006. Forecasting and decision theory. In: Elliott, C.G.G., Timmermann, A. (Eds.), Handbook of Economic Forecasting, vol. 1. Elsevier, pp. 81–98.
Le Cam, L., 1986. Asymptotic Methods in Statistical Decision Theory. Springer-Verlag, New York.
Lehmann, E., Romano, J., 2005. Testing Statistical Hypotheses. Springer.
Manganelli, S., 2009. Forecasting with judgment. J. Bus. Econ. Stat. 27 (4), 553–563.
Manganelli, S., 2023. Double conditioning: the hidden connection between Bayesian and classical statistics. ECB Working Paper Series 2786. pp. 1–24.
Manski, C., 2013. Public Policy in an Uncertain World: Analysis and Decisions. Harvard University Press.
Manski, C.F., 2021. Econometrics for decision making: building foundations sketched by Haavelmo and Wald. Econometrica 89 (6), 2827–2853.
Markowitz, H., 1952. Portfolio selection. J. Finance 39, 47–61.
Newey, W., McFadden, D., 1994. Large sample estimation and hypothesis testing. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, vol. 4. North Holland, Amsterdam.
Patton, A., Timmermann, A., 2012. Forecast rationality tests based on multi-horizon bounds. J. Bus. Econ. Stat. 30 (1), 1–17.
Savage, L., 1954. The Foundations of Statistics. John Wiley & Sons, New York.
Stein, C., 1956. Inadmissibility of the usual estimator for the mean of a multivariate distribution. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability 1, pp. 197–206.
Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. Science 185 (4157), 1124–1131.
Wald, A., 1950. Statistical Decision Functions. John Wiley & Sons, New York.