Deciding with Judgment

Simone Manganelli*
European Central Bank
simone.manganelli@ecb.int

March, 2018

Abstract

Non sample information is hidden in frequentist statistics in the choice of the hypothesis to be tested and of the confidence level. Explicit treatment of these elements provides the connection between Bayesian and frequentist statistics. A frequentist decision maker starts from a judgmental decision and moves to the closest boundary of the confidence interval of the first order conditions, for a given loss function. This statistical decision rule does not perform worse than the judgmental decision with a probability equal to the confidence level. For any given prior, there is a mapping from the sample realization to the confidence level which makes Bayesian and frequentist decision rules equivalent. Frequentist decision rules can be interpreted as decisions under ambiguity.

*I would like to thank for useful comments and suggestions Joel Sobel, Harald Uhlig, Andrew Patton, Bruno Biais, Frank Diebold, participants at the 2016 NBER Summer Institute workshop on Forecasting and Empirical Methods in Macro and Finance, the 2016 Duke/UNC conference on New Developments in Measuring and Forecasting Financial Volatility, the 2017 Southampton Finance and Econometrics Workshop, the 2017 NBER-NSF Seminar on Bayesian Inference in Econometrics and Statistics in St. Louis, the Berlin conference Modern Econometrics Faces Machine Learning, the 10^{th} SoFiE conference in New York, the inaugural conference of the HeiKaMEtrics Network on Financial Econometrics in Heidelberg, the meeting of the German Statistical Association in Rostock, as well as seminar participants at the ECB, Humboldt University Berlin, Technische Universitaet Dresden, the University of Bern, the UZH Finance Seminar in Zurich, Tinbergen Institute's Econometrics seminar in Amsterdam, the Finance Seminar Series at the Goethe University in Frankfurt am Main, and the de Finetti Risk Seminars at Bocconi University in Milan.

Keywords: Statistical Decision Theory; Statistical Risk Aversion; Portfolio Selection.
JEL Codes: C1; C11; C12; C13; D81.

1 Introduction

Most people take decisions in an uncertain environment without resorting to formal statistical analysis. I refer to these decisions as judgmental decisions. Statistical decision theory uses data to prescribe optimal choices under a set of assumptions, but has not explicit role for judgmental decisions. This paper is concerned with the following questions. Is a given judgmental decision optimal? If not, how can it be improved in a precise statistical sense?

For concreteness, consider an investor who is about to take the decision $\tilde{a}$, say, to hold all her assets in cash. She asks an econometrician for advice on whether she should invest some of her money in a stock market index. The best prediction of the econometrician depends on an estimated parameter $\hat{\theta}$, which is affected by estimation uncertainty. For a given utility function provided by the investor, the econometrician can construct a loss function $L(\theta, \tilde{a})$, the loss experienced by the investor if the decision $\tilde{a}$ is taken and the true parameter is θ. Suppose the econometrician is able to recover the distribution of the gradient $\nabla_a L(\hat{\theta}, \tilde{a})$ around the true, but unknown θ. It is possible to test whether the investor's decision $\tilde{a}$ is optimal by testing the null hypothesis that $\nabla_a L(\theta, \tilde{a})$ is equal to zero. If the null hypothesis is not rejected, the econometrician cannot recommend any deviation from $\tilde{a}$. If the null hypothesis is rejected, statistical evidence suggests that marginal deviations from $\tilde{a}$ decrease the loss function relative to $L(\theta, \tilde{a})$.

Denote with α the confidence level used to implement the hypothesis testing. The investor is facing the decision problem depicted in figure 1. The investor has two possible choices. She can hold on to her judgmental decision $\tilde{a}$, denoted by the action J, incurring in the loss $L(\theta, \tilde{a})$. Alternatively, she can follow the econometrician's advice, which is equivalent to accepting the bet $\mathcal{L}_\alpha$. In this case, she does not know whether she is facing the upper part of the decision tree, denoted by the node H_0, or the lower part, denoted by H_1. H_0 is the bad scenario, in which the null hypothesis is true, so that any deviation from the judgmental decision $\tilde{a}$ results in a higher loss. A marginal $\varepsilon > 0$ move away from $\tilde{a}$ results in the loss $L(\theta, \tilde{a}) + |\nabla_a L(\theta, \tilde{a})|\varepsilon$. H_1 is the favorable scenario, as one correctly rejects the null hypothesis that

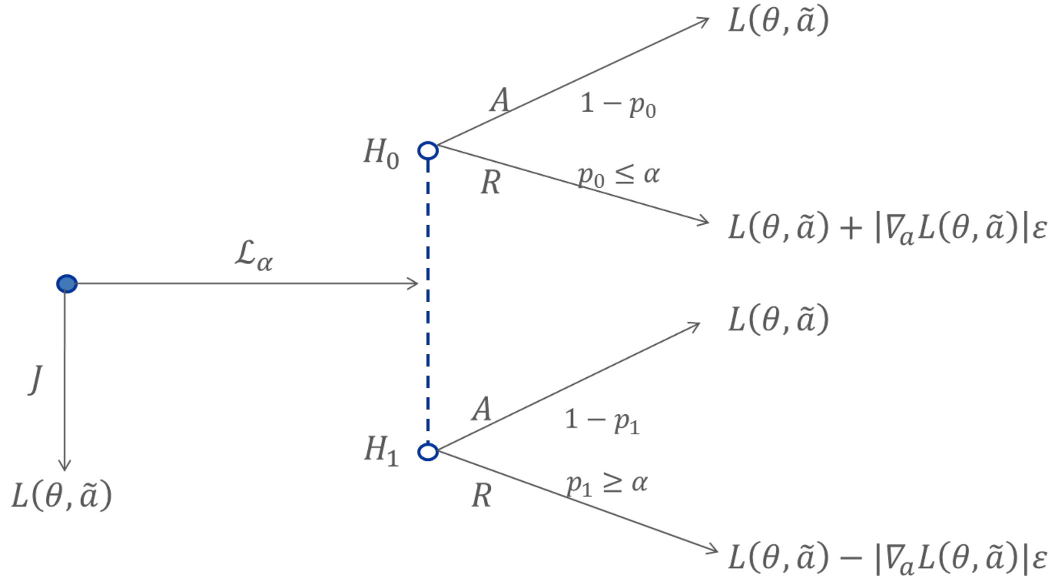


Figure 1: Statistical Decision Tree

$\tilde{a}$ is optimal, producing decisions with lower loss. In this case, a marginal ε move away from $\tilde{a}$ results in the loss $L(\theta, \tilde{a}) - |\nabla_a L(\theta, \tilde{a})|\varepsilon$. 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 confidence level α, which puts an upper bound to the probability that the null is wrongly rejected when it is true. Notice that α represents also the lower bound probability of correctly rejecting H_0 when it is false.

In case of rejection, the investor faces a new, but identical decision problem, except that $\tilde{a}$ is replaced by $\tilde{a} \pm \varepsilon$ (the sign depends on the sign of the empirical gradient). This new action will be rejected if $\nabla_a L(\hat{\theta}, \tilde{a} \pm \varepsilon)$ also falls in the rejection region. Iterating this argument forward, the preferred decision of the investor is the action $\tilde{a} \pm \Delta$ which lies at the boundary of the $(1-\alpha)$-confidence interval of $\nabla_a L(\hat{\theta}, \tilde{a} \pm \Delta)$, the point where the null hypothesis that the decision $\tilde{a} \pm \Delta$ is optimal can no longer be rejected. This decision is characterized by the fact that it will produce a higher loss than the original judgmental decision $\tilde{a}$ with probability at most α. In other words, a decision maker may prefer to abandon her judgmental decision $\tilde{a}$ (the action J) and follow a statistical procedure (the bet $\mathcal{L}_\alpha$), if this produces a worse decision only with probability at most α.

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). Recent contributions in decision theory focus on modeling behavior when beliefs cannot be quantified by a unique Bayesian prior. See Gilboa and Marinacci (2013) and Marinacci (2015) for comprehensive reviews. This paper, however, is not concerned with the axiomatic foundations of decision theory, but rather with how data can be used to help decision makers take better decisions. It falls within Clive Granger's tradition that *'to obtain any kind of best value for a point forecast, one requires a criterion against which various alternatives can be judged'* (Granger and Newbold, 1986, p. 121). This tradition has been duly continued by, among many, Patton and Timmermann (2007), Patton and Timmermann (2012) and Elliott and Timmermann (2016). Papers on forecasting using Bayesian statistical decision theory include Chamberlain (2000) and Geweke and Whiteman (2006). Manski has published influential contributions on the use of statistical decision theory in the presence of ambiguity for partial identification of treatment response (Manski 2000, Manski 2004, Manski 2013, Dominitz and Manski 2017)

The paper is structured as follows. Section 2 sets up the decision environment and introduces the concept of judgment in frequentist statistics. The judgment is defined as a pair formed by a judgmental decision $\tilde{a}$ and a confidence level α associated with it. Judgment is used to set up the hypothesis to test whether the action $\tilde{a}$ is optimal. Two key results of this section are that the frequentist decision rule with judgment is admissible, and that it is either the judgmental decision itself or is at the boundary of the confidence interval of the sample gradient of the loss function. The admissibility result holds generally also for the maximum likelihood estimate, thus solving Stein paradox. The key element behind this result is that the frequentist decision rule incorporating judgment conditions on the observed sample realization.

Section 3 discusses the choice of the confidence level α. This is a mapping from the *p-value* of the first order conditions evaluated at the judgmental decision into the unit interval. The optimal frequentist decision rule has the feature that it does not perform worse than the judgmental decision with probability at most equal to α. Since the confidence level controls the probability of Type I errors, and these in turn determine whether the statistical decision performs worse than the judgmental decision, α can be interpreted as the willingness of the decision maker to take statistical risk. This concept is closely linked to the idea of ambiguity aversion. The section also discusses how the confidence level α can be elicited with a simple experiment involving urns à la Ellsberg.

Section 4 establishes the equivalence between Bayesian and frequentist decisions. To understand the link with Bayesian decision rules, consider that since the confidence level determines the width of the confidence interval and the optimal frequentist decision moves from the judgmental decision to the closest boundary of the confidence interval, the confidence level determines the deviation from the judgmental decision. This deviation is zero if the confidence level is equal to the *p-value* and it is maximum if the confidence level is equal to one, which implies a zero width confidence interval and coincides with the maximum likelihood decision. By choosing different mappings from the *p-value* to the confidence level, it is possible to choose different confidence intervals and therefore generate any convex combination between the judgmental and the maximum likelihood decisions. This provides the bridge to establish the equivalence between Bayesian and frequentist decision rules: For a given sample realization and for any Bayesian posterior distribution shrinking from the prior to the maximum likelihood decisions, there is a mapping from the *p-value* to the confidence level which shrinks from the judgmental to the maximum likelihood decisions by exactly the same amount.

Section 5 uses an asset allocation problem as a working example to illustrate the empirical performance of various decision rules. The decision problem is a simple asset allocation of an investor who holds €100 and has to decide how much to invest in an Exchange Trading Fund replicating the EuroStoxx50 index. Section 6 briefly concludes.

2 Statistical Decision Rules

Judgmental information is usually incorporated in statistical analysis in the form of prior distributions and exploited via the Bayesian updating. This section introduces the concept of judgment in a frequentist context and shows how hypothesis testing can be used to arrive at optimal frequentist decisions.

For concreteness, consider an asset allocation problem of an investor who is holding all her wealth in cash and has to decide what fraction $a \in \mathbb{R}$ of her wealth to invest in a stock market index. Let $X \sim N(\theta, 1)$ denote the stock market index return, with known variance, but unknown expected return equal to the parameter $\theta \in \mathbb{R}$, and suppose that one realization x of the random variable is available. Let also the return on cash be zero. Suppose the investor wants to minimize a mean-variance loss function à la Markowitz (1952), $L(\theta, a) \equiv -a\theta + 0.5a^2$. The main object of interest of the analysis will be the first derivative evaluated at the maximum likelihood estimator $\hat{\theta}(X) = X$, that is $\nabla_a L(\hat{\theta}(X), a) = -X + a$. Notice that for this specific loss function $\nabla_a L(\hat{\theta}(X), a) - \nabla_a L(\theta, a) \sim N(0, 1)$. I formally define the decision environment as follows.

Definition 2.1 (**Decision Environment**). *The decision environment is characterized by the following elements:*

1. $X - \theta \sim \Phi(x)$*, for* $\theta \in \mathbb{R}$*, where* $\Phi(x)$ *is the cdf of the standard normal distribution.*
2. *The sample realization* $x \in \mathbb{R}$ *is observed.*
3. $a \in \mathbb{R}$ *denotes the action of the decision maker.*
4. *The decision maker has loss function* $L(\theta, a) = -a\theta + 0.5a^2$.

Remark: General case — This example is more general than it may seem.[1] Let $\mathbf{X}_t \in \mathbb{R}^q$, $q \in \mathbb{N}$, $t = 1, 2, ...$, be a vector of random variables, with cdf $F_t^\theta(\mathbf{x}_t)$, where $\theta \in \mathbb{R}^p$, $p \in \mathbb{N}$, is a vector of unknown statistical parameters. Let $\hat{\theta}_n(\mathbf{X}) \in \mathbb{R}^p$ be an extremum estimator (in the sense of Newey and Powell, 1994) which depends on $\mathbf{X} \equiv [\mathbf{X}_1, ..., \mathbf{X}_n]'$, a matrix of dimensions (n, q), with $n \in \mathbb{N}$ large enough to allow for asymptotic approximation. Assume that the sample $\mathbf{x} \equiv [\mathbf{x}_1, ..., \mathbf{x}_n]'$ is observed and let $\tilde{\mathbf{b}} \in \mathbb{R}^p$ be a nonrandom vector (an exogenous judgmental decision, to be defined later in this section). The decision variable $a \in \mathbb{R}$ determines the decision on the parameter estimate via the following relationship: $b(a) = a\tilde{\mathbf{b}} + (1 - a)\hat{\theta}_n(\mathbf{x})$. Finally, assume that the loss function $L(\theta, a)$ is differentiable and strictly convex. Using a mean value expansion of the first order conditions around the population parameter θ:

$$\nabla_a L(\hat{\theta}_n(\mathbf{X}), a) = \nabla_a L(\theta, a) + \nabla_{a\theta} L(\bar{\theta}_n(\mathbf{X}), a)(\hat{\theta}_n(\mathbf{X}) - \theta)$$

[1] See also section 5 of Manganelli (2009) for a similar generalization.

where $\bar{\theta}_n(\mathbf{X})$ lies between $\hat{\theta}_n(\mathbf{X})$ and θ. Assuming standard regularity conditions are satisfied (see for instance Newey and Powell, 1994, or White, 1994) so that $\sqrt{n}(\hat{\theta}_n(\mathbf{X})-\theta) \overset{d}{\sim} N(0,\Sigma)$, the statistic $\sqrt{n}\hat{\sigma}^{-1}(\nabla_a L(\hat{\theta}_n(\mathbf{X}),a)-\nabla_a L(\theta,a)) \overset{d}{\sim} N(0,1)$, where the term $\hat{\sigma}$ is a consistent estimate of the asymptotic variance $\sigma^2 \equiv \nabla'_{a\theta}\ L(\theta,a)\Sigma\ \nabla_{a\theta}L(\theta,a)$. We are therefore in a situation identical to the one represented by the decision environment of Definition 2.1, except for the fact that the gradient may now be a nonlinear function of a. However, the strict convexity assumption guarantees that there is a one to one mapping between a and the gradient. $\square$

Before discussing the alternative decisions, consider the following standard definition of a decision rule (Wald, 1950):

Definition 2.2 (**Decision Rule**). $\delta(X) : \mathbb{R} \to \mathbb{R}$ *is a decision rule, such that if* $X = x$ *is the sample realization,* $\delta(x)$ *is the action that will be taken.*

2.1 The Bayesian Decision

The Bayesian solution assumes that the decision maker uses subjective information in the form of a prior distribution over the unknown parameter θ.

Definition 2.3 (**Prior**). *The subjective information of the decision maker is summarized by the prior cdf* $\mu(\theta)$ *over the parameter* $\theta \in \mathbb{R}$.

Once the prior information is specified, the optimal Bayesian decision minimizes the expected loss function, using the posterior distribution to compute the expectation.

Bayesian Decision — *Consider the decision environment of Definition 2.1. If the decision maker knows her prior distribution* $\mu(\theta)$*, the Bayesian decision is:*

$$\delta^{\mu}(x) = \arg\min_a \int (-a\theta + 0.5a^2)d\mu(\theta|x)$$

where $\mu(\theta|x)$ *denotes the posterior distribution.*

The Bayesian decision has the considerable merit of having an axiomatic justification, being grounded in a decision theoretic framework.[2] It has, how-

[2]The axiomatic foundation of the Bayesian approach goes back to the works of Ramsey, De Finetti and Savage. See Gilboa (2009) and Gilboa and Marinacci (2013) for recent surveys of the literature.

ever, been criticized along two main points. First, it implies lack of uncertainty aversion, while the experimental evidence built on Ellsberg's paradox suggests otherwise. Starting with Schmeidler (1989) and Gilboa and Schmedler (1989), the literature has developed alternative axioms which account for ambiguity aversion: Agents do not optimize over a single known prior, but they minimize the maximum loss with respect to a prespecified set of priors.

A second criticism, leveled against both strands of literatures, is that neither the prior distribution nor the set of prior distributions over which the optimization is performed, are known. Without the specification of priors, none of the above procedures is applicable. In fact, lack of knowledge of the priors was one of the motivations to develop classical statistics.[3]

2.2 The Frequentist Decision

Classical statistics as developed by Neyman and Fisher has no explicit role for *epistemic* uncertainty (as defined by Marinacci, 2015), as it was motivated by the desire for objectivity. Non sample information is, nevertheless, implicitly introduced in various forms, in particular in the choice of the confidence level and the choice of the hypothesis to be tested. This subsection shows how to make explicit the non sample information hidden in the frequentist approach. Explicit treatment of non sample information in the classical paradigm provides a connection between Bayesian and frequentist statistics.

2.2.1 Judgment

In a frequentist setting, there is no prior distribution to help arriving at a decision. One solution often used is the plug-in estimator, which replaces the unknown θ parameter with its sample counterpart (see, for instance, chapter 3 of Elliott and Timmermann, 2016). In the decision environment of Definition 2.1, this is given by x. Once θ is replaced by x, the optimal decision is $\delta^{Plug-in}(x) = x$. The problem with such decision is that it does

[3]Incidentally, until the first half of the 20^{th} century, the term classical statistics was referring to the Bayesian approach. Neyman and Fisher, the fathers of frequentist statistics, had sharp scientific disagreements, but were united in their skepticism of using the Bayesian framework for practical problems: *"When a priori probabilities are not available (which [Fisher] presumed to be always the case and which I agree is almost always the case), then the formula of Bayes is not applicable"* (Neyman, 1952, p. 193).

not minimize the loss function, but rather its sample equivalent, and neglects estimation error.

The solution proposed in this paper is to explicitly incorporate subjective information of the decision maker also in a frequentist setting and use hypothesis testing to arrive at a decision. Subjective information takes the form of judgment, defined as follows:

Definition 2.4 (**Judgment**). *The subjective information of the decision maker is summarized by the pair $A \equiv \{\tilde{a}, \alpha\}$, which is referred to as **judgment**. $\tilde{a} \in \mathbb{R}$ is the **judgmental decision**. $\alpha \in [0, 1]$ 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 confidence level (usually 1%, 5% or 10%). I say nothing about how the judgment is formed. It is a primitive to the decision problem, like the loss function and the Bayesian priors. Note that for any prior $\mu(\theta)$, there is a judgmental decision $\tilde{a}_\mu$ which is observationally equivalent to a Bayesian decision with no posterior updating: $\tilde{a}_\mu$ solves the problem $\min_a \int L(\theta, a) d\mu(\theta)$.

α reflects the confidence that the decision maker has in her judgmental decision and determines the amount of statistical evidence needed to abandon $\tilde{a}$. The choice of α is also closely linked to the choice of prior distributions and determines the frequentist decision, after observing the realization $X = x$. The decision in the light of the statistical evidence x is taken on the basis of hypothesis testing, which is where I turn next.

2.2.2 Hypothesis Testing

Given the judgmental decision $\tilde{a}$, the decision maker can test whether $\tilde{a}$ is optimal by testing if the gradient $\nabla_a L(\theta, \tilde{a}) = -\theta + \tilde{a}$ is equal to zero. A test statistic for the gradient can be obtained by replacing θ with its maximum likelihood estimator X.

The exact hypothesis to be tested depends on the sample realization x. Suppose without loss of generality that the empirical gradient $\nabla_a L(x, \tilde{a}) = -x + \tilde{a}$ is negative. This implies that values of a higher than $\tilde{a}$ decrease the empirical loss function. The decision maker is interested, however, in the population value of the loss function. If the population gradient is positive, higher values of a would increase the loss function, rather than decrease it.

The null hypothesis to be tested is therefore that the population gradient has opposite sign relative to the sample gradient.

Two cases are possible:

i) $\underline{-x + \tilde{a} \leq 0}$

$$H_0 : -\theta + \tilde{a} \geq 0 \quad \text{vs} \quad H_1 : -\theta + \tilde{a} < 0 \tag{1}$$

ii) $\underline{-x + \tilde{a} \geq 0}$

$$H_0 : -\theta + \tilde{a} \leq 0 \quad \text{vs} \quad H_1 : -\theta + \tilde{a} > 0 \tag{2}$$

As in any hypothesis testing procedure, the decision maker can make two types of errors. She can wrongly reject the null hypothesis (Type I error). This occurs with probability $p_0 \leq \alpha$, the confidence level chosen by the decision maker. The economic interpretation of this type of error is that although the decision maker decreases the sample approximation of the loss function by moving from $\tilde{a}$ towards x, in fact she increases the loss function in population. Alternatively, she can fail to reject the null hypothesis when it is false (Type II error). This happens with probability $1 - p_1$, where $p_1 \geq \alpha$ is the power of the test. The economic interpretation in this state of the world is that the decision maker could have decreased her loss function in population, but statistical uncertainty prevented her from doing so. The trade-off is well known: A small α generally implies also a small power for values of $\tilde{a}$ close to θ. Therefore, a smaller probability of Type I errors results in a greater probability of Type II errors. It is up to the preferences of the decision maker to decide how to solve this trade-off. The reasoning is summarized in table 1, which is the reduced form representation of the bet $\mathcal{L}_\alpha$ in the statistical decision tree of figure 1.

2.2.3 Decision

In an hypothesis testing decision problem, only two actions are possible: The null hypothesis is either accepted or rejected. Given the confidence level α, it is possible to define the rejection and acceptance regions to arrive at a decision. Let $0 \leq \gamma \leq 1$ and $\Phi(c_\alpha) = \alpha$, and consider again the two cases, conditional on the sample realization x. Given the judgment $A = \{\tilde{a}, \alpha\}$, the test functions $\psi_i^A(x), i = \{1, 2\}$ associated with the hypotheses (1)-(2) are:

Table 1: Hypothesis testing

		Decision	
		A	R
Truth	H_0	Avoid higher loss $1 - p_0$	Higher loss $p_0 \leq \alpha$
	H_1	Fail to lower loss $1 - p_1$	Lower loss $p_1 \geq \alpha$

Note: The null hypothesis tests whether the gradient of the loss function has opposite sign with respect to the sample gradient. The alternative hypothesis is that population and sample gradients have the same sign.
α and p_1 are the size and power of the test.

i) $\underline{-x + \tilde{a} \leq 0}$

$$\psi_1^A(x) = \begin{cases} 0 & \text{if } c_{\alpha/2} < -x + \tilde{a} \leq 0 \\ \gamma & \text{if } -x + \tilde{a} = c_{\alpha/2} \\ 1 & \text{if } -x + \tilde{a} < c_{\alpha/2} \end{cases} \tag{3}$$

ii) $\underline{-x + \tilde{a} \geq 0}$

$$\psi_2^A(x) = \begin{cases} 0 & \text{if } 0 \leq c_{1-\alpha/2} < -x + \tilde{a} \\ \gamma & \text{if } -x + \tilde{a} = c_{1-\alpha/2} \\ 1 & \text{if } -x + \tilde{a} > c_{1-\alpha/2} \end{cases} \tag{4}$$

The test function determines whether the null hypothesis is rejected or not. The null hypothesis is a statement about the population gradient evaluated at the judgmental decision $\tilde{a}$. It says that *marginal* moves from $\tilde{a}$ towards the *observed* maximum likelihood estimate x increase the loss function. If it is not rejected at the given confidence level α, the chosen action is $\tilde{a}$. Rejection of the null hypothesis, on the other hand, prescribes to marginally move from $\tilde{a}$ towards x. Suppose that $-x+\tilde{a}<0$ and let's denote the new action marginally away from $\tilde{a}$ with $a_\varepsilon=\tilde{a}+\varepsilon$, for $\varepsilon>0$ and sufficiently small. Notice that a_ε is not random and it is possible to test whether it is optimal, by testing again whether additional marginal moves from a_ε towards x increase the loss function. Iterative application of this hypothesis testing procedure delivers the following frequentist decision:

Theorem 2.1. ***(Frequentist decision)*** *Consider the decision environment of Definition 2.1. If the decision maker has judgment $A=\{\tilde{a},\alpha\}$, the frequentist decision is:*

$$\delta^A(X)=\begin{cases}\tilde{a}(1-\psi_1^A(X))+(x+c_{\alpha/2})\psi_1^A(X) & \text{if } -x+\tilde{a}\leq 0\\ \tilde{a}(1-\psi_2^A(X))+(x+c_{1-\alpha/2})\psi_2^A(X) & \text{if } -x+\tilde{a}\geq 0\end{cases} \tag{5}$$

where $c_\alpha\equiv\Phi^{-1}(\alpha)$ and $\psi_i^A(X), i=\{1,2\}$ are the test functions defined in (3)-(4).

Proof — See Appendix.

The proof of the theorem clarifies the intuition behind this result. For a given confidence level α, rejecting the initial null hypothesis (1) or (2) automatically implies rejection of all null hypotheses for $a\in(\tilde{a},x+c_{\alpha/2})$ or $a\in(x+c_{1-\alpha/2},\tilde{a})$, i.e. until the closest boundary of the $(1-\alpha)$ confidence interval is reached. That is the action for which the gradient of the loss function is no longer statistically different from zero.

The next theorem shows that the frequentist decision is optimal.

Theorem 2.2. ***(Admissibility)*** *The decision $\delta^A(X)$ of Theorem 2.1 is admissible.*

Proof — See Appendix.

The admissibility result is a direct consequence of Karlin-Rubin theorem applied to the test functions (3)-(4), which is itself an extension of the celebrated Neyman-Pearson lemma. An implication of Theorem 2.2 is that it solves Stein paradox, as illustrated in the following remark.

Remark: Stein paradox is no longer a paradox — Stein (1956) considered the following example. Let $\mathbf{Y}=[Y_1,...,Y_p]'$ be independent random variables with $Y_i\sim N(\vartheta_i,1)$, $i=1,...,p$, and assume that a single draw

$\mathbf{y} = [y_1, ..., y_p]'$ is observed for each of them. He showed that the maximum likelihood estimator $\hat{\vartheta}(\mathbf{Y}) = \mathbf{Y}$ is not admissible with a quadratic loss function $L(\vartheta, \hat{\vartheta}(\mathbf{Y})) = 0.5||\vartheta - \hat{\vartheta}(\mathbf{Y})||^2$ when $p \geq 3$. Stein's example can be reformulated in the context of this paper as follows. Define the action $\mathbf{b}(a) \equiv a\tilde{\mathbf{b}} + (1-a)\mathbf{y}$, where $a \in \mathbb{R}$, $\tilde{\mathbf{b}} \in \mathbb{R}^p$ is some non stochastic vector and $\mathbf{y}$ is the sample realization. The quadratic loss function can be expressed as a function of the decision variable a, $L(\vartheta, a) = 0.5\Sigma_{i=1}^{p}(\vartheta_i - b_i(a))^2$, where $b_i(a)$ denotes the i^{th} element of the p-vector $\mathbf{b}(a)$. Also, any given judgment $A = \{\tilde{\mathbf{b}}, \alpha\}$ can be equivalently expressed as $A = \{\tilde{a}, \alpha\}$, where $\tilde{a} = 1$. The gradient of the loss function evaluated at $\hat{\vartheta}(\mathbf{Y}) = \mathbf{Y}$ is $\nabla_a L(\mathbf{Y}, \tilde{a}) = \Sigma_{i=1}^{p}(Y_i - b_i(\tilde{a}))(y_i - \tilde{b}_i)$, which is normally distributed with mean $\mu(a) = \Sigma_{i=1}^{p}(\vartheta_i - b_i(a))(y_i - \tilde{b}_i)$ and variance $\sigma^2 = \Sigma_{i=1}^{p}(y_i - \tilde{b}_i)^2$. It is therefore possible to test whether the judgmental decision $\tilde{a}$ is optimal, using the same procedure illustrated in the previous subsections. Suppose that $\nabla_a L(\mathbf{y}, \tilde{a}) \leq 0$, so that the hypothesis to be tested is $H_0 : \nabla_a L(\vartheta, \tilde{a}) \geq 0$. The maximum likelihood estimate can be recovered by choosing a confidence level $\alpha = 1$, as this implies that the test statistic will always fall in the rejection region and the frequentist decision is $\delta^A(\mathbf{Y}) = \hat{a}$, where $\hat{a}$ is such that $\nabla_a L(\mathbf{y}, \hat{a}) = 0$.[4] In Stein's example this is given by $\hat{a} = 0$. Since the conditions for the admissibility of $\delta^A(X)$ in Theorem 2.2 do not exclude $\alpha = 1$, it is possible to conclude that the decision associated with the maximum likelihood estimate is admissible also for $p \geq 3$.

This result is in fact quite obvious if one realizes that the maximum likelihood estimate (as opposed to the estimator) is not random. Any action which is always equal to a fixed quantity is admissible, since it has lowest risk when the optimal action coincides with that fixed quantity. The paradox is resolved by considering the expression for $\delta^A(X)$ in Theorem 2.1. When $\alpha = 1$, it follows that $c_{\alpha/2} = 0$ and $\psi_i(X) = 1$, $\forall X$, which in turn implies that the optimal decision $\delta^A(X) = x$ coincides with the maximum likelihood estimate. The formula conditions on the observed value x, which is nonrandom. Stein's formulation of the frequentist decision does not condition on the sample realization, but treats x as a random variable. It is this lack of conditioning — that is the replacement of the sample realization x with its random variable X in the formula for $\delta^A(X)$ — that generates the paradox. $\square$

[4] See the next section for a formal justification of this result.

3 Choosing the Confidence Level

Careful choice of the confidence level α generates the most common estimators in econometrics as a special case of the decision rule $\delta^A(X)$. A key insight of this section is that although α is chosen before seeing the data and remains a primitive of the decision problem, its choice is conditional on the realization of the random variable x.

The choice of the confidence level α may generally be considered as a mapping into the interval $[0,1]$ from the *p-value* of the test statistic $-X+\tilde{a}$ that determines the test functions (3)-(4) under $H_0 : \nabla_a L(\theta,\tilde{a}) = -\theta+\tilde{a} = 0$.

Definition 3.1 (**Choice of the Confidence Level**). *The confidence level of the decision maker is:*

$$\alpha|x = g(\tilde{\alpha}) : (0,1] \to [0,1]$$

where

$$\tilde{\alpha} \equiv \begin{cases} 2\Phi(-x+\tilde{a}) & \textit{if } -x+\tilde{a} \leq 0 \\ 2(1-\Phi(-x+\tilde{a})) & \textit{if } -x+\tilde{a} \geq 0 \end{cases}$$

Since the *p-value* $\tilde{\alpha}$ is determined by the sample realization x, the choice of α is conditional on x. I have made explicit this fact with the notation $\alpha|x$.

Any judgmental decision $\tilde{a}$ with *p-value* $\tilde{\alpha} \geq \alpha$ implies that the decision taken is the judgmental decision itself, because in this case $-x+\tilde{a}$ falls within the confidence interval of the decision rule (5). One can therefore impose the condition that $g(\tilde{\alpha}) \geq \tilde{\alpha}$ without affecting the decision rule (5). This condition ensures that decisions are always at the boundary of the confidence interval. Note that in this case the decision rule (5) simplifies to:

$$\delta^{A|x}(x) = \begin{cases} x + c_{\alpha/2|x} & \textit{if } -x+\tilde{a} \leq 0 \\ x + c_{1-\alpha/2|x} & \textit{if } -x+\tilde{a} \geq 0 \end{cases} \tag{6}$$

after imposing $\psi_i^A(x) = 1$, $i = \{1,2\}$. The notation $\delta^{A|x}$ is equivalent to δ^A, but makes explicit the conditioning on x. When computing the risk function of $\delta^{A|x}(X)$ in (5), the expectation is taken with respect to the test functions $\psi_i^A(X), i = \{1,2\}$, but the confidence level α will continue to be conditioned on the observed sample realization x.

Here are some common examples of how the function $g(\tilde{\alpha})$ in Definition 3.1 is chosen:

1. Maximum likelihood:

$$\alpha|x = 1, \forall\ \tilde{\alpha} \in (0, 1]$$

2. Pretest estimator with threshold $\bar{\alpha}$:

$$\alpha|x = \begin{cases} 1 & \text{if } \tilde{\alpha} \in (0, \bar{\alpha}) \\ \tilde{\alpha} & \text{if } \tilde{\alpha} \in [\bar{\alpha}, 1] \end{cases}$$

3. Subjective classical estimator with threshold $\bar{\alpha}$ (Manganelli, 2009):

$$\alpha|x = \begin{cases} \bar{\alpha} & \text{if } \tilde{\alpha} \in (0, \bar{\alpha}) \\ \tilde{\alpha} & \text{if } \tilde{\alpha} \in [\bar{\alpha}, 1] \end{cases}$$

4. Judgmental decision:

$$\alpha|x = \tilde{\alpha}, \forall\ \tilde{\alpha} \in (0, 1]$$

It is easy to verify that each of these estimators is obtained by replacing the corresponding choice of $g(\tilde{\alpha})$ in the decision rule (6). The maximum likelihood estimator always disregards any judgmental decision, by setting the confidence level equal to 1. In this case, $c_{\alpha/2|x} = c_{1-\alpha/2|x} = 0$ and $\delta^{A|x}(x) = x$. The pretest estimator maintains the confidence level $\tilde{\alpha}$ if the test statistic falls within the confidence interval determined by $\bar{\alpha}$, but it is increased to 1 otherwise. The subjective classical estimator maintains the threshold probability $\bar{\alpha}$ for *p-values* lower than $\bar{\alpha}$, otherwise it is equal to $\tilde{\alpha}$. The judgmental decision coincides in this case with an unconstrained minimax decision rule, which never abandons the judgmental decision, by setting the confidence level always equal to $\tilde{\alpha}$.

3.1 Interpreting the confidence level as statistical risk aversion

The confidence level α has an intuitive economic interpretation provided by the following theorem.

Theorem 3.1 (**Economic interpretation of the confidence level**). *Consider the decision environment of Definition 2.1 and assume the decision maker has judgment $A|x \equiv (\tilde{a}, \alpha|x)$. The decision rule $\delta^{A|x}(X)$ in (5) performs worse than the judgmental decision $\tilde{a}$ with probability not greater than $\alpha|x$:*

$$P_\theta(L(\theta, \delta^{A|x}(X)) > L(\theta, \tilde{a})) \leq \alpha|x \tag{7}$$

Proof — See Appendix.

For a given judgmental decision $\tilde{a}$ and sample realization x, moving away from $\tilde{a}$ will either increase or decrease the loss function. The decision maker can attach a probability to these events by performing the following frequentist thought experiment: Draw an infinite amount of samples $\{x^i\}_{i=1}^{\infty}$ from $X \sim N(\theta, 1)$ and compute the fraction of times the statistical decision rule $\delta^{A|x}(x^i)$ performs worse than the judgmental decision. This probability will depend on the unknown population value of θ, but Theorem 3.1 shows that it is bounded from above by the chosen confidence level. It formalizes the intuition that a decision maker may be willing to abandon her judgmental decision and follow a statistical procedure only if there are sufficient guarantees that it does not result in an action that is worse than the judgmental decision. Since statistical procedures are subject to randomness, the decision maker cannot be sure that this will be the case and the guarantee can only be expressed in terms of a probability.

Theorem 3.1 suggests an alternative interpretation of the confidence level. Given that $\alpha|x$ represents the upper bound of the probability that the statistical decision rule performs worse than the judgmental decision, the confidence level reflects the willingness of the decision maker to take *statistical risk*. Notice that this concept is distinct from the standard concept of risk aversion, as summarized by the weight given to the portfolio variance in the loss function $L(\theta, a)$ in the decision environment of Definition 2.1.

The idea of statistical risk aversion is closely linked to the concept of ambiguity (or uncertainty) aversion. The decision tree of figure 1 can be used to define the decision maker's attitude towards statistical risk. Classical statistics imposes $\alpha = 1$, which is equivalent to assuming that the decision maker is *statistical risk lover*, that is $\mathcal{L}_1 \succ J$. An *extreme statistical risk averse* decision maker, on the other hand, always prefers sure outcomes to situations involving uncertainty, that is $J \succ \mathcal{L}_\alpha$, $\forall \alpha \in (0, 1]$. Intermediate degrees of *statistical risk aversion* can be represented by appropriate choices of the confidence level α, such that $\mathcal{L}_\alpha \succ J$.

The decision tree of figure 1 can also be used to give a heuristic proof that the optimal decision must be at the boundary of the confidence interval. Given the statistical risk preference α, $\tilde{a}$ cannot be optimal if $\tilde{\alpha} < \alpha$, because $\mathcal{L}_\alpha \succ \mathcal{L}_{\tilde{\alpha}}$. Suppose that $-x + \tilde{a} \leq 0$ and the preferred action is $a_\Delta = x + c_{\alpha/2|x} + \Delta$. If $\Delta < 0$, $\alpha_\Delta \equiv P(-X + a_\Delta < -x + a_\Delta | \theta = a_\Delta) < \alpha$. Since $\mathcal{L}_\alpha \succ \mathcal{L}_{\alpha_\Delta}$, it is preferred to marginally move away from a_Δ. If $\Delta > 0$,

$\alpha_\Delta > \alpha$. This implies that action $a_{\Delta-\varepsilon}$ was preferred to action a_0 for $\varepsilon > 0$ sufficiently small, which is a contradiction because $\mathcal{L}_\alpha \succ \mathcal{L}_{\alpha_{\Delta-\varepsilon}}$.

3.2 Eliciting the Confidence Level

One possible strategy to elicit the degree of statistical risk aversion (i.e., the confidence level α) could be to run an experiment à la Ellsberg (1961) where the decision maker can choose among different couples of urns. Consider two urns with 100 balls each. Urn 1 contains only white and black balls, Urn 2 contains white and red balls. If the black ball is extracted, the respondent loses €100. If the red ball is extracted, the respondent wins an amount in euros which gives an increase in utility equivalent to the reduction in utility produced by the loss of €100. If the white ball is extracted, nothing happens. The respondent can choose among the composition of the urns described in table 2.

The respondent faces uncertainty, as she does not know whether the balls are drawn from Urn 1 or Urn 2 and their exact composition. By accepting one of the bets from 0 to 100, she can control the upper bound probability of losing in case balls are drawn from Urn 1. By choosing this upper bound probability, she automatically chooses the lower bound probability of winning in case the ball is drawn from Urn 2. An extreme statistical risk averse player would always choose bet 0, which is equivalent to not participating to the bet and holding on to the judgmental decision $\tilde{a}$. A statistical risk loving player would choose bet 100. In general, players with higher degrees of statistical risk aversion would choose bets with lower numbers.

4 Equivalence between Bayesian and Frequentist Decisions

Careful choice of the confidence level $\alpha|x$ allows one to arrive at frequentist decisions which are equivalent to Bayesian decisions.

Let's start by noticing that the decision rule $\delta^{A|x}(x)$ of equation (5) is a shrinkage estimate.

Proposition 4.1 (**Skrinkage**). *Given the judgment $A|x \equiv (\tilde{a}, \alpha|x)$, for any sample realization x, the decision rule $\delta^{A|x}(x)$ of Theorem 2.1 is a shrink-*

Table 2: Experiment to elicit the confidence level α

	Urn 1		Urn 2	
Bet	White	Black	White	Red
0	100	0	100	0
1	≥ 99	≤ 1	≤ 99	≥ 1
2	≥ 98	≤ 2	≤ 98	≥ 2
...	...	...	...	...
98	≥ 2	≤ 98	≤ 2	≥ 98
99	≥ 1	≤ 99	≤ 1	≥ 99
100	0	100	0	100

Note: The decision maker can choose one of the bets from 0 to 100. By accepting the bet, she will face Urn 1 or Urn 2 without knowing the probability with which the urn is chosen. If a white ball is extracted, nothing happens. If a black ball is extracted, the decision maker loses €100. If a red ball is extracted, she wins a utility equivalent euro amount. In accepting the bet, the decision maker can partially choose the composition of the urns. For instance, by choosing bet 2, she knows that Urn 1 does not contain more than 2 black balls and Urn 2 contains at least 2 red balls. A statistical risk loving decision maker chooses bet 100. An extreme statistical risk averse decision maker chooses bet 0. Decision makers with higher degrees of statistical risk aversion choose bets with lower numbers.

age estimate of the type $\delta^{A|x}(x) = (1-h)x + h\tilde{a}$*, where* $h \equiv c_{\alpha/2|x}/c_{\tilde{\alpha}/2} = c_{1-\alpha/2|x}/c_{1-\tilde{\alpha}/2} \in [0,1]$.

Proof — See Appendix.

This decision rule is a convex combination of the judgmental decision and the maximum likelihood estimate. The amount of shrinkage is determined by the factor h, which is a combination of data (as represented by x) and judgmental information (as represented by the judgmental decision $\tilde{a}$ and the associated confidence level $\alpha|x$). Note the similarity with Bayesian estimators of a model with unknown mean, known variance, and Normal prior (see example 1, p.127 of Berger 1985). Let the prior be a Normal density $N(\tilde{a}, \tau^2)$. Since $X \sim N(\theta, 1)$, the posterior mean is $(\tau^2/(1+\tau^2))x + (1/(1+\tau^2))\tilde{a}$. Informative priors (that is, low τ^2) imply a posterior mean close to the prior mean. In Proposition 4.1, judgment with low confidence level (that is, an $\alpha|x$ close to $\tilde{\alpha}$) implies h close to 1 and therefore a decision which is close to the judgmental decision $\tilde{a}$. On the other hand, uninformative priors (that is, large τ^2) imply a posterior mean close to the maximum likelihood estimator, which in turn is equivalent to judgment with high confidence level (that is, an $\alpha|x$ close to 1 and h close to 0).

The following Theorem shows that for any Bayesian decision associated with a given prior there is a corresponding choice of $\alpha|x = g(\tilde{\alpha})$ which produces an equivalent frequentist decision.

Theorem 4.1 (**Equivalence between Bayesian and Frequentist Decisions**). *Consider the decision environment of Definition 2.1. For any prior distribution* $\mu(\theta)$ *such that* $\tilde{a} = \arg\min_a \int(-a\theta + 0.5a^2)\ d\mu(\theta)$ *and the Bayesian decision* $\delta^\mu(x)$ *lies between* $\tilde{a}$ *and* x*, the Bayesian decision is equivalent to the frequentist decision* $\delta^{A|x}(x)$*, when*

$$\alpha|x = \begin{cases} 2\Phi(\delta^\mu(x) - x) & \text{if } -x + \tilde{a} \leq 0 \\ 2(1 - \Phi(\delta^\mu(x) - x)) & \text{if } -x + \tilde{a} \geq 0 \end{cases} \tag{8}$$

Proof — See Appendix.

I consider here the comparison with two special Bayesian estimators, which have been analyzed at length by Magnus (2002) in the case $\tilde{a} = 0$.

Bayesian estimator based on Normal prior — Assuming that the prior over the parameter θ is Normally distributed with mean zero and variance $1/c$, the optimal Bayesian decision is:

$$\delta^N(x) = (1+c)^{-1}x \tag{9}$$

Bayesian estimator based on Laplace prior — If the prior over the parameter θ is distributed as a Laplace with mean zero and scale parameter c, the optimal Bayesian decision is:

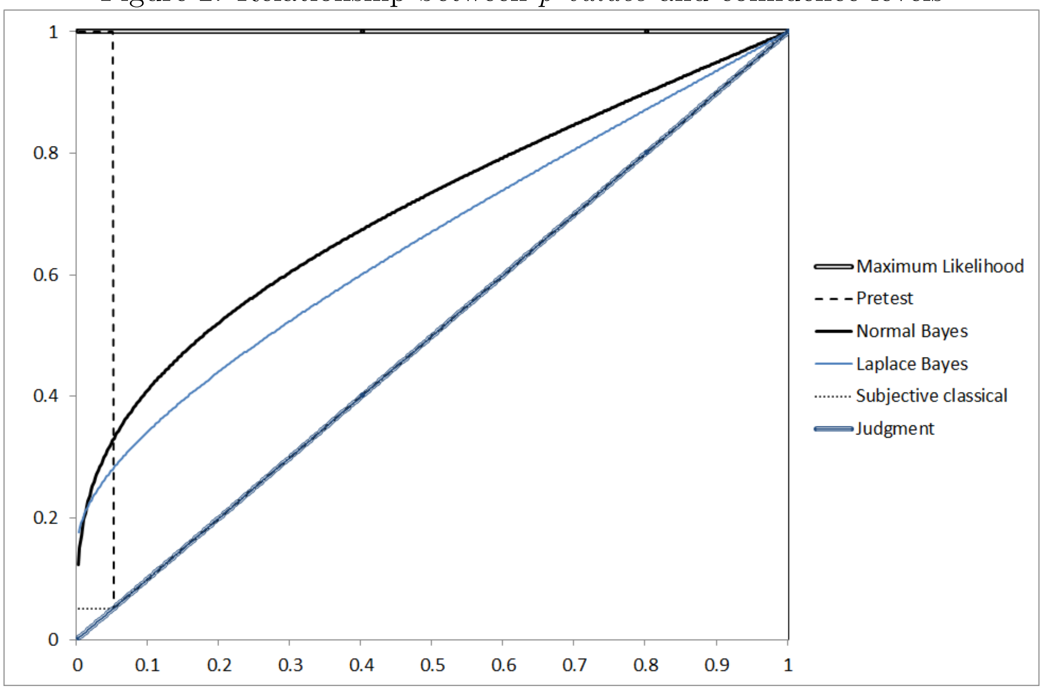
$$\delta^L(x) = x - c \cdot \frac{1 - \exp(2cx)\frac{\Phi(-x-c)}{\Phi(x-c)}}{1 + \exp(2cx)\frac{\Phi(-x-c)}{\Phi(x-c)}} \tag{10}$$

In figure 2, I compare in the space $(\alpha|x, \tilde{\alpha})$ the confidence levels associated with the decision rules discussed in section 3 and the two Bayesian decision rules above. The mapping $\alpha|x = g(\tilde{\alpha})$ for the Bayesian decisions is obtained by substituting (9) and (10) in (8), and considering that there is a one to one mapping between x and $\tilde{\alpha}$, which is the *p-value* associated with the test statistic $-x + \tilde{\alpha}$ (conditional on $-x + \tilde{\alpha}$ being positive or negative).

Note how all statistical decision rules have a confidence level mapping which falls between the two extreme decision behaviors: the judgmental decision (which corresponds to the decision of an extreme statistical risk averse decision maker, where no data is taken into consideration) and the maximum likelihood estimator (which corresponds to the decision of a statistical risk loving decision maker, where no judgment is taken into consideration). The judgmental decision is described by the diagonal line in the space $(\alpha|x, \tilde{\alpha})$. As already discussed in section 3, any point below this diagonal line is equivalent to its vertical projection on the diagonal, in the sense that they all imply that the frequentist decision (5) coincides with the judgmental decision $\tilde{\alpha}$. The confidence level $\alpha|x$ associated with the maximum likelihood estimator, instead, does not depend on $\tilde{\alpha}$ and is always equal to 1.

The confidence level of the pretest estimator is equal to that of the judgmental decision for intermediate values of $\tilde{\alpha}$, but jumps discontinuously to the maximum likelihood for extreme values of $\tilde{\alpha}$ (less than 10% in the example of figure 2). It has the feature that small changes in $\tilde{\alpha}$ may trigger abrupt changes in the confidence level. The choice of the confidence level of the subjective classical estimator proposed by Manganelli (2009) avoids the discontinuity of the pretest estimator.

Figure 2: Relationship between *p-values* and confidence levels



Note: The horizontal axis reports the *p-value* (α) of the gradient $\nabla_a L(\theta, a)$ evaluated at $\theta = x$ and $\tilde{a} = 0$, the judgmental decision. The vertical axis is the chosen confidence level $\alpha | x = g(\tilde{\alpha})$. The figure plots the mapping corresponding to six alternative estimators. Pretest and subjective classical estimators (Manganelli, 2009) are based on 10% confidence levels. The Normal and Laplace Bayesian estimators are based on priors with zero mean and unit variance.

The figure reports also the confidence levels associated with the two Bayesian estimators. The plot reveals a few interesting features.

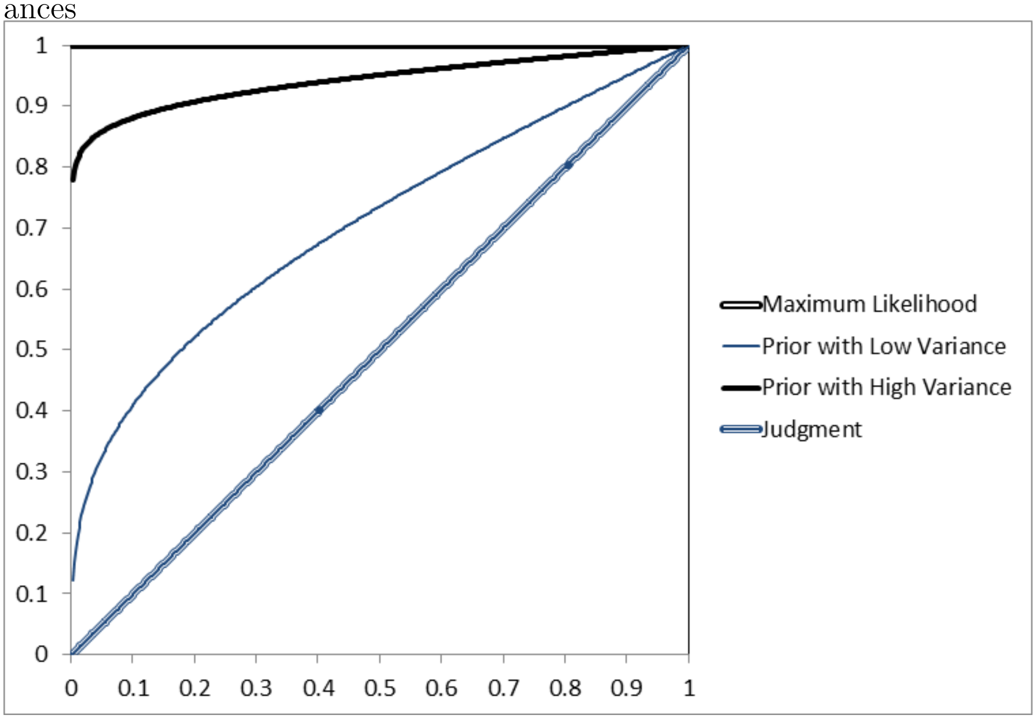
First, the figure shows that the confidence levels associated with the Normal Bayesian estimator converges to zero as $\tilde{\alpha}$ goes to zero. It shrinks relatively less when the initial judgment is extremely bad, an odd choice of the statistical risk preference.

Second, Bayesian econometrics requires the decision maker to express her judgment on the statistical parameters of the random variables, rather than on the decision variables directly. The whole literature on prior elicitation notwithstanding, choosing priors is often a formidable task, and, as already mentioned, it was one of the main motivations driving Neyman and Fisher to develop frequentist statistics. In the context of the asset allocation problem discussed in this paper, two prior distributions with same mean and standard deviation can lead to very different portfolio allocations. Asking whether her prior distribution of the mean has fat or thin tails strikes me as putting an unrealistic burden on the decision maker. If one leaves the unconditional, univariate domain, the requests in terms of prior specification become even more challenging.

Third, this paper shows that imposing priors on parameters is equivalent to imposing specific statistical risk preferences on the decision maker. Consider the case in which the decision maker is a central banker who has to decide the level of interest rates. The Bayesian approach requires central bankers to express their priors for the parameters of the macro-econometric model of the economy. Even though there is by now a rich literature on Bayesian estimation of Dynamic Stochastic General Equilibrium models (see for instance Smets and Wouters 2007 and subsequent applications), it is my impression that the decision making body of a central bank has little clue about the construction of these models, let alone the multivariate priors of the underlying parameters. It is usually the expert who imposes priors to arrive at some reasonable estimate of the model. Econometricians and decision makers should be aware that this is not an innocuous exercise and that it has direct implications on the willingness of the central banker to tolerate statistical risk in her decision process.

Figure 3 reports the confidence level mappings associated with the Bayesian decision with Normal priors of different precision, one with high variance and another one with low variance. The estimator with low variance attaches greater weight to the prior and results in a confidence level mapping closer to the judgmental decision. In the limit, as the variance goes to zero, the

Figure 3: Confidence level mappings for Normal priors with different variances



Note: The horizontal axis reports the *p-value* (α) of the gradient $\nabla_a L(\theta, a)$ evaluated at $\theta = x$ and $\tilde{a} = 0$, the judgmental decision. The vertical axis is the chosen confidence level $\alpha|x = g(\tilde{\alpha})$. The figure plots the confidence level mappings corresponding to Bayesian decisions based on Normal priors with mean zero and different variances.

confidence level mapping of the Normal Bayesian decision converges to the mapping of the judgmental decision. At the other extreme, as the variance of the prior goes to infinity, the Bayesian estimator attaches lower weight to the prior and its confidence level mapping converges to the one of the maximum likelihood estimator.

The same figure can be given the interpretation of a Normal Bayesian decision with different sample sizes. As the sample size increases, the Bayesian decision attaches less weight to the prior and more weight to the data. Asymptotically, the confidence level mapping of the Bayesian decision converges to the one of the maximum likelihood estimator. This represents an additional idiosyncrasy of the Bayesian approach: The Bayesian posterior updating imposes time varying statistical risk preferences on the decision maker.[5]

4.1 Equivalence with Ambiguity Averse Decisions

The previous discussion has highlighted a serious shortcoming of Bayesian decisions, namely that the choice of priors imposes specific, sample dependent statistical risk preferences on the decision maker. More standard criticism of the Bayesian approach focuses on two other issues, which are linked to the broader issue of prior robustness: It requires the decision maker to know the prior distribution and imposes that the decision maker is not averse to uncertainty.

Figure 2 presents a simple example of how prior robustness may be a real issue. The two Bayesian estimators are based on prior distributions which have been calibrated to have both zero mean and unit variance. One key difference is that the Laplace priors, unlike the Normal distribution, has fat tails. As already shown by the risk analysis of Magnus (2002) and Manganelli (2009), Bayesian estimators based on apparently 'close' priors can have very different properties. The issue of prior robustness is well-known and acknowledged in the literature. It is known, for instance, that

[5] In its extreme interpretation, the Bayesian approach requires that states of the world resolve all uncertainty, so that each individual chooses her strategy only once at the beginning of their lifetime. See the discussion in section 2.4 of Gilboa and Marinacci (2013). This is clearly unrealistic. Still, inconsistencies of static theories of ambiguity when extended to dynamic decision problems and in particular the lack of a useful notion of updating mechanism are a major source of controversy among Bayesians (see, for instance, Al-Najjar and Weinstein, 2009, and the follow up comment by Siniscalchi, 2009).

one of main sources of nonrobustness in Bayesian estimation is linked to the thickness of the tails of the prior distributions (see the discussion at page 197 in Berger 1985). The example 2, p. 111, of Berger (1985) raises similar issues by comparing decisions based on Normal and Cauchy priors matched to have the same median and interquartiles.

Both issues of prior robustness and ambiguity aversion may be partly addressed by considering classes of priors, instead of a single prior. Gilboa and Schmeidler (1989) have shown that an ambiguity averse decision maker characterized by a set of priors Γ minimizes the expected loss using the worst possible prior from the set Γ. Cerreia Vioglio et al. (2013a and 2013b) provide an axiomatic characterization which clarifies the relationship between the maxmin approach of decision theory under ambiguity and the minimax approach of robust statistics in the presence of parametric prior uncertainty. Formally:

Definition 4.1 (**Set of Priors**). *The subjective information of the decision maker is summarized by the set of priors* $\Gamma \equiv \{\mu(\theta), \theta \in \Theta \subset \mathbb{R}\}$.

Ambiguity Averse Decision — *Consider the decision environment of Definition 2.1. If the decision maker knows her set of prior distributions* Γ, *the ambiguity averse decision is:*

$$\delta^{\Gamma}(x) = \arg\min_{a} \max_{\mu\in\Gamma} \int (-a\theta + 0.5a^2) d\mu(\theta|x)$$

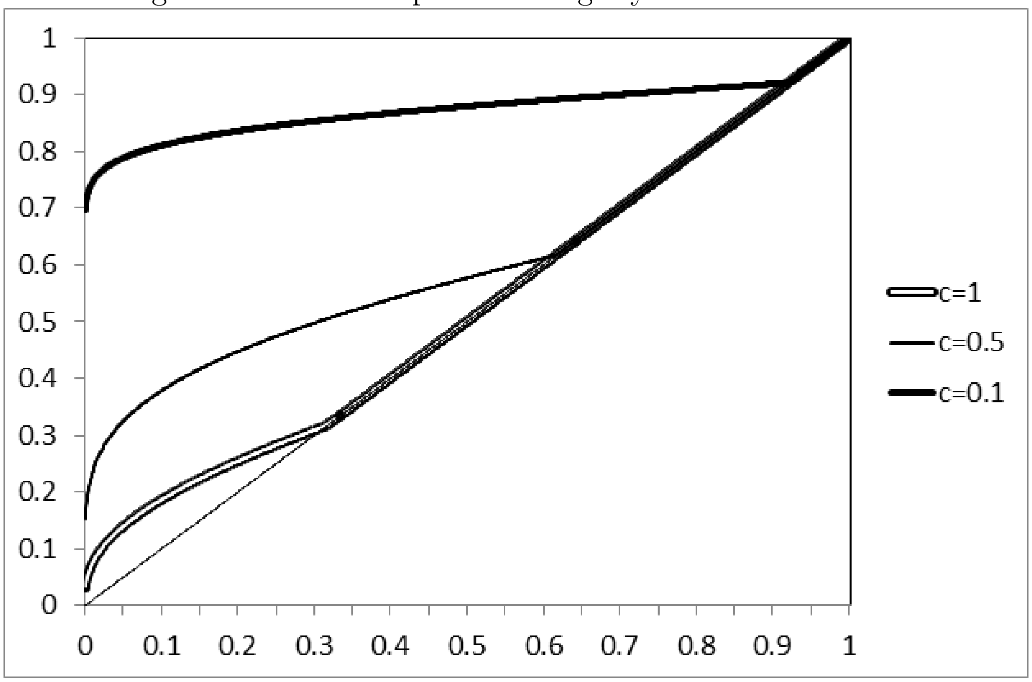
where $\mu(\theta|x)$ *denotes the posterior distribution for a given* $\mu \in \Gamma$.

The frequentist decision (5) is flexible enough to cover also the case of ambiguity aversion. The equivalence between frequentist decision and ambiguity averse decision can be proven in a similar way to that of Theorem 4.1. For each sample realization x, there is a confidence level mapping $\alpha|x$ which sets $\delta^{\Gamma}(x) = \delta^{A|x}(x)$. It may be helpful to illustrate this equivalence by considering a specific example.

Consider again the decision environment of Definition 2.1 and let $\Gamma = \{\mu : \mu \text{ is } N(\pi, c^{-1}), \underline{\pi} \leq \pi \leq \bar{\pi}\}$. Since the posterior mean of θ is given by $(1+c)^{-1}(x + c\pi)$ and the loss function is linear in θ, it must be:

$$\max_{\mu\in\Gamma} \int (-a\theta + 0.5a^2) dF^{\mu}(\theta|x) =$$
$$= -a(\underline{\pi} I(a > 0) + \bar{\pi} I(a < 0))(1+c)^{-1}c - a(1+c)^{-1}x + 0.5a^2$$

Figure 4: Relationship with ambiguity averse decision rule



Note: The horizontal axis reports the *p-value* (α) of the gradient $\nabla_a L(\theta, a)$ evaluated at $\theta = x$ and $\tilde{a} = 0$, the judgmental decision. The vertical axis is the chosen confidence level $\alpha|x = g(\tilde{\alpha})$. The figure plots the confidence level mappings corresponding to an ambiguity averse decision maker, who chooses from the set of priors $\Gamma = \{\mu : \mu \text{ is } N(\pi, c^{-1}), \underline{\pi} \leq \pi \leq \bar{pi}\}$, for $\underline{\pi} = -1$, $\bar{\pi} = 1$, and different levels of precision c.

where $F^{\mu}(\theta|x)$ denotes the posterior updating of a given μ distribution in Γ.

The ambiguity averse decision associated with the set of priors Γ is obtained by setting the first derivative with respect to a of the above expression equal to 0:

$$\delta^{\Gamma}(x) = \begin{cases} (1+c)^{-1}(x + c\underline{\pi}) & \text{if } x > -c\underline{\pi} \\ 0 & \text{if } -c\bar{\pi} \leq x \leq -c\underline{\pi} \\ (1+c)^{-1}(x + c\bar{\pi}) & \text{if } x < -c\bar{\pi} \end{cases} \tag{11}$$

The confidence level associated with decision (11) is reported in figure 4 for $\underline{\pi} = -1$, $\bar{\pi} = 1$ and $c = \{1, 0.5, 0.1\}$. The figure reveals a link between the subjective classical estimator of Manganelli (2009) and ambiguity aversion. If the sample realization falls within the interval $(-c\bar{\pi}, -c\underline{\pi})$, the ambiguity averse investor chooses not to enter the stock market. This is like the situation of the frequentist decision rule (5) when the test function $\psi_i^A(x) = 0$, which occurs when the test statistic $-x + \tilde{a}$ falls within the frequentist confidence interval.

The link between frequentist confidence intervals and Knightian uncertainty was first suggested by Bewley (1988), who showed how classical confidence regions correspond to sets of posterior means derived from a standardized set of prior distributions.[6] He, however, did not formulate the frequentist non sample information in the form of judgment as done in this paper, which makes the use of frequentist procedures both practical and theoretically sound.

A comparison of figures 3 and 4 also illustrates how Bayesian decisions do not take uncertainty into account. Looking at the Bayesian confidence level of figure 3, it is clear that $\alpha|x > \tilde{\alpha}\ \forall\tilde{\alpha}$, which in turn implies that the null hypothesis that the judgmental decision $\tilde{a}$ is optimal is always rejected. For the subjective classical decision of Manganelli (2009) and the ambiguity averse decision (11) this is not the case.

5 Empirical evidence

Section 4 highlighted the statistical differences among the estimators. An equally important question is whether the estimators produce portfolio allocations with significant economic differences. I address this issue by bringing the estimators to the data, solving a standard 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, it is well-known that plug-in estimators of the portfolio weights produce volatile asset allocations which usually perform poorly out of sample, due to estimation errors. Bayesian approaches offer no better performance. There is a vast literature documenting the empirical failures of the mean-variance model and suggesting possible fixes (Jobson and Korkie 1981, Brandt 2007). DeMiguel, Garlappi and Uppal (2009), however, provide evidence that none of the existing solutions consistently outperforms a simple, non-statistically driven equal weight portfolio allocation. Other

[6]Bewley's work on decision under uncertainty has an interesting history, as told by Gilboa (2009), footnotes 127 and 128. Bewley's (1988) working paper was later published as Bewley (2011).

examples of non-statistically driven portfolios could be that of an investment manager with some benchmark against which she is evaluated, or that of a private household who may have all her savings in a bank account (and therefore a zero weight in the risky investment). DeMiguel et al. (2009) raise an important point: How to improve on a given judgmental allocation? The framework of this paper offers an answer to this question: For a given sample realization and confidence level, the frequentist decision rules will not perform worse than the given judgmental allocation with a probability equal to the confidence level.

To implement the statistical decision rules, I take a monthly series of closing prices for the EuroStoxx50 index, from January 1999 until December 2015. EuroStoxx50 covers the 50 leading Blue-chip stocks for the Eurozone. The data is taken from Bloomberg. The closing prices are converted into period log returns. Table 3 reports summary statistics.

Table 3: Summary statistics

Obs	Mean	Std. Dev.	Median	Min	Max	Jarque Bera
206	-0.06%	5.57%	0.66%	-20.62%	13.70%	0.0032

Note: Summary statistics of the monthly returns of the EuroStoxx50 index from January 1999 to December 2015. The Jarque Bera statistic is the *p-value* of the null hypothesis that the time series is normally distributed.

The exercise consists of forecasting the next period optimal investment in the Eurostoxx50 index of a person who holds €100 cash. I take the first 7 years of data as pre-sample observations, to estimate the optimal investment for January 2006. The estimation window then expands by one observation at a time, the new allocation is estimated, and the whole exercise is repeated until the end of the sample.

To directly apply the decision rules discussed in the previous sections, which assume the variance to be known, I transform the data as follows. I first divide the return series of each window by the full sample standard deviation, and next multiply them by the square root of the number of observations in the estimation sample. Denoting by $\{\tilde{x}_t\}_{t=1}^n$ the original time series of log returns, let σ be the full sample standard deviation and $n_1 < n$ the size of the first estimation sample. Then, for each $n_1 + s$, $s = 0, 1, 2, ..., n - n_1 - 1$, define:

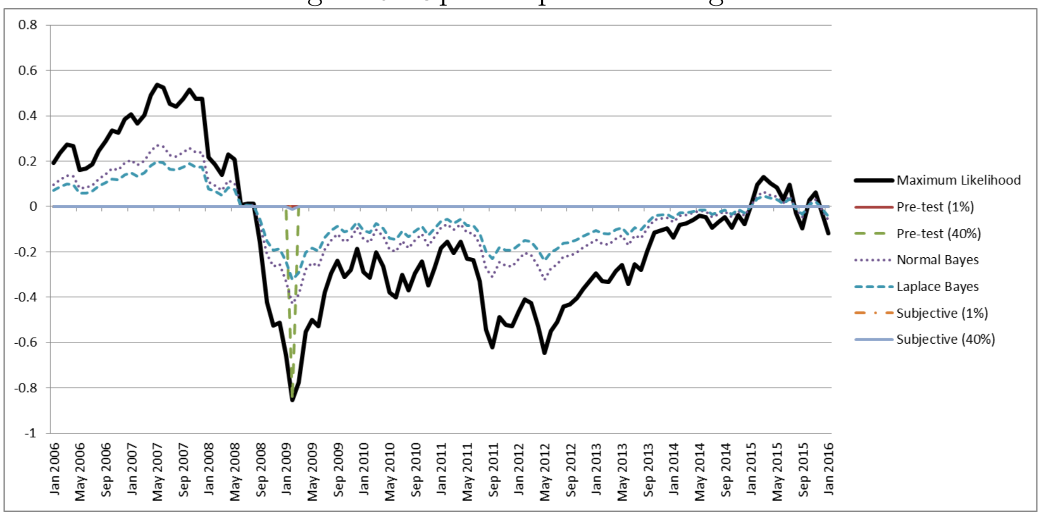
$$\{x_t\}_{t=1}^{n_1+s} \equiv \{\sqrt{(n_1+s)}\tilde{x}_t/\sigma\}_{t=1}^{n_1+s} \quad \text{and} \quad \bar{x}_{n_1+s} \equiv (n_1+s)^{-1}\sum_{t=1}^{n_1+s} x_t \tag{12}$$

I 'help' the estimates by providing the full sample standard deviation, so that the only parameter to be estimated is the mean return. Under the assumption that the full sample standard deviation is the population value, by the central limit theorem $\bar{x}_{n_1+s}$ is normally distributed with variance equal to one and unknown mean. We can therefore implement the decision rules discussed in the preceding sections of the paper, using the single observation $\bar{x}_{n_1+s}$ for each period n_1+s.

The results of this exercise are reported in figures 5 and 6. Figure 5 plots the optimal weights obtained from the different decision rules. A few things are worth noticing. First, the weight associated with the maximum likelihood decision is the most volatile, as it is the one that suffers the most from estimation error. The Bayesian decision rules are shrunk towards zero, the one based on a Normal prior being shrunk less than the one based on Laplace prior, consistently with the pattern shown in figure 2. Pretest and subjective classical decision rules predict an optimal weight equal to zero, as the *p-value* is almost always greater than the chosen threshold $\bar{\alpha}$: The data is just too noisy to suggest a departure from the judgmental decision. One needs to increase the threshold to 40% for this to be the case. That is the spike observed in February 2009 for the pretest decision, which for that month coincides with the maximum likelihood decision (remember that when the *p-value* is less than $\bar{\alpha}$, the pretest decision rule reverts to the maximum likelihood). The weight associated with the subjective classical decision rule with 40% confidence threshold exhibits just a small blip, as it goes to the boundary of the confidence interval associated with α instead of moving all the way to the maximum likelihood.

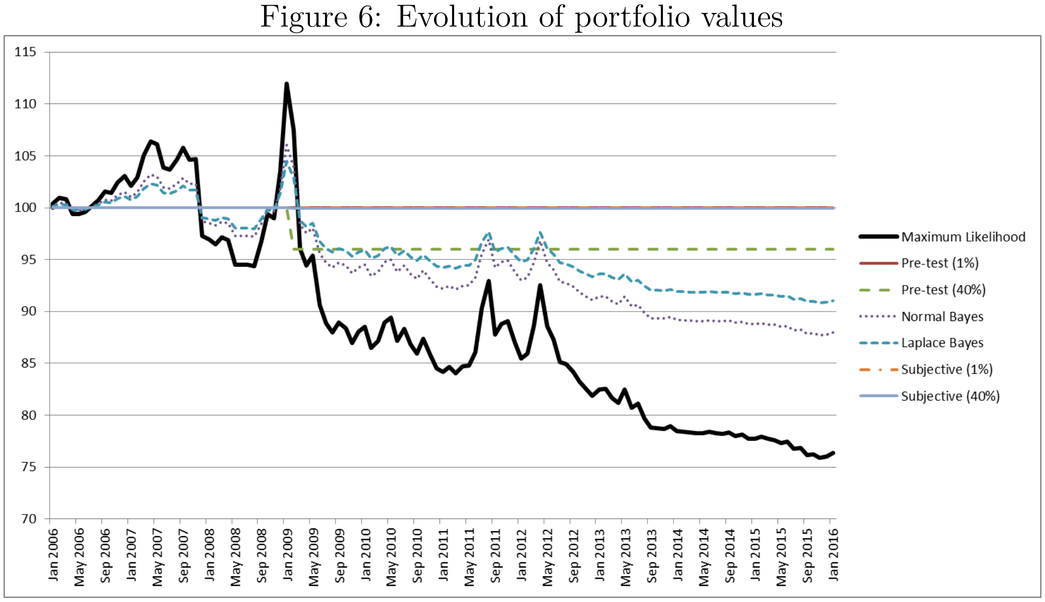
Figure 6 reports the portfolio values associated with the strategy of an investor who would re-optimize each month and decide how much to allocate in the EuroStoxx50 index on the basis of the decision rules associated with different confidence thresholds. Suppose the starting value of the portfolio in January 2006 is €100. By the end of the sample, after 10 years, an investor using the maximum likelihood decision rule would have lost one quarter of the value of her portfolio. The situation is slightly better with the Bayesian decision rules, as they imply a loss of between 9% and 12%. The pretest decision rule with threshold of 40% would have lost little less than

Figure 5: Optimal portfolio weights



Note: Optimal weights according to the different decision rules of an investor choosing between cash and the EuroStoxx50 index. Weights are re-estimated each month by expanding the estimation window by one data point. The first 7 years — from January 1999 until December 2005 — are used to produce the first estimate in January 2006.

Figure 6: Evolution of portfolio values



Note: Time evolution of the value of a portfolio invested in cash and the EuroStoxx50 index following the investment recommendations of the different decision rules.

5%. Note that the entire loss comes from shorting the position and following the predictions of the maximum likelihood decision in February 2009. In all the other months there is no investment in the stock market. The other three decision rules – the pretest with 1% and the subjective classical with confidence thresholds at 1% and 40% – do not lose anything because they never predict deviating from the judgmental allocation of holding all the money in cash.[7] In fact, the subjective classical decision rule with confidence threshold of 40% does lose something, as like the pretest decision it rejects the judgmental allocation in February 2009. However, unlike the pretest decision which goes to the boundary of the 0% confidence interval (the maximum likelihood), the subjective classical decision only moves to the boundary of the 60% confidence interval, so that the overall losses are contained to less than 1% and barely visible from the chart.

The point of this discussion is not to evaluate whether one decision rule is better than the other, as the decision rules differ only with respect to the choice of the confidence level, which is a subjective choice like the choice of the loss function. The purpose is rather to illustrate the implications of choosing different confidence levels. By choosing the maximum likelihood estimator, one has no control on the statistical risk she is going to bear. With the subjective classical estimator, instead, the investor chooses a constant probability of underperforming the judgmental allocation: she can be sure that the resulting asset allocation is not worse than the judgmental allocation with the chosen probability. The case of the EuroStoxx50, however, represents only one possible draw, which turned out to be particularly adverse to the maximum likelihood and Bayesian estimators. Had the resulting allocation implied positive returns by the end of the sample, maximum likelihood and Bayesian estimators would have outperformed the subjective classical estimators. There is no free lunch: decision rules with lower confidence thresholds produce allocations with greater protection to underperformance relative to the judgmental allocation, but also have lower upside potential. In statistical jargon, lower confidence levels protect the decision maker from Type I errors, but imply higher probabilities of Type II errors.

I illustrate this intuition with a simulation. I generate several sets of 500 random samples of 206 observations using the empirical distribution of the

[7] This finding is consistent with results from the literature on portfolio choice under ambiguity, which shows that there exists an interval of prices within which no trade occurs. See for instance Guidolin and Rinaldi (2013) and the references therein.

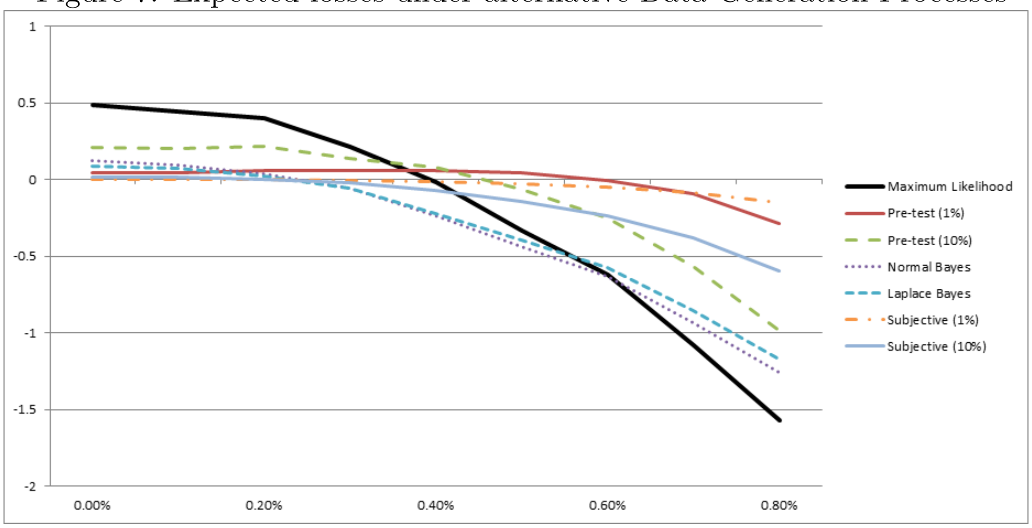
EuroStoxx50 time series from January 1999 until December 2015. Each set is generated by adding different means to the empirical distribution, starting from zero (which would be the equivalent of replicating EuroStoxx50 500 times, after subtracting its empirical mean) and then progressively increasing it, so that the zero judgmental allocation becomes less and less accurate. I then replicate the same estimation strategy used to produce the results in figures 5 and 6, i.e. I use the first 85 observations (the equivalent of 7 years of data) to estimate the optimal allocation and increase the sample one observation at a time to estimate the next period allocation. This exercise is repeated for all random samples, 500 of them, and for each of the different means. The results are reported in figures 7 and 8.

Figure 7 plots the average expected loss associated with each estimator against the different means simulated in the exercise. Remember that the judgmental decision implies zero allocation in the risky asset, which would be the optimal allocation when the population mean is equal to zero. As we move to the right of the horizontal axis, we are therefore considering data generation processes which are less and less in line with the judgmental allocation. Since I know the data generation process, I can compute the population expected loss. For values of the mean close to zero, the subjective classical estimators dominate all the others, the one with 1% confidence thresholds being better than the one with 10% confidence thresholds for smaller values of the mean. As the population mean increases beyond 0.3% the Bayesian estimators start to perform better than the subjective classical estimators. It is only when the population mean exceeds 0.6% that the maximum likelihood estimator starts to dominate the others. Not surprisingly, decisions based on higher confidence levels generate relatively lower expected loss only when the judgmental allocation is far from the optimal one, as can be seen by the Normal Bayesian estimator dominating the Laplace Bayesian one for values of the mean greater than 0.3%. To paraphrase a famous quote by Clive Granger, investors with good judgment do better than investors with no judgment, who do better than investors with bad judgment.[8]

Figure 8 shows the percentage of times the statistical rule does worse than the judgmental allocation. It reports the unconditional percentage of times (out of the 500 replications) that the various estimators underperform

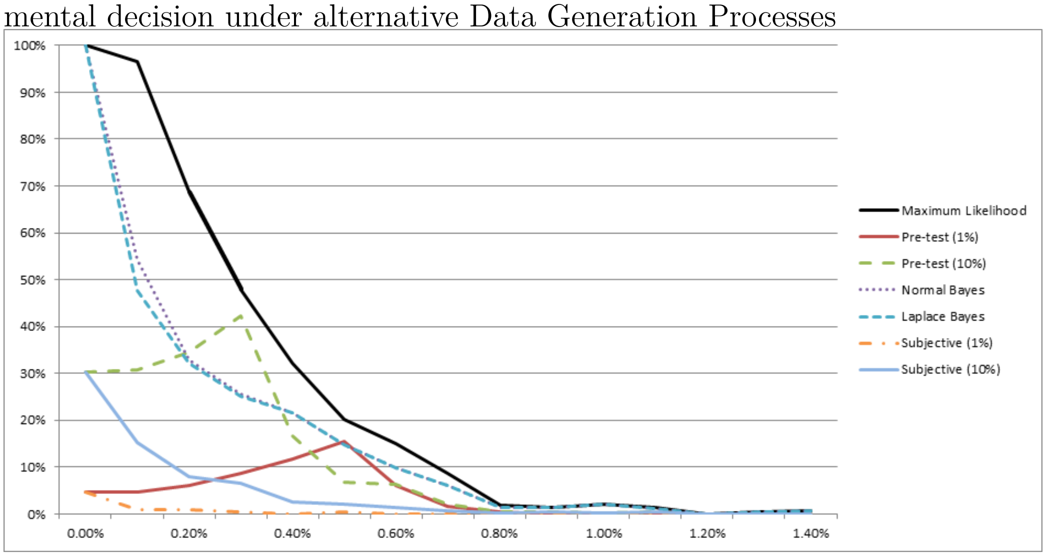
[8]The original quote is '*a good Bayesian... is better than a non-Bayesian. And a bad Bayesian... is worse than a non-Bayesian*' (see Phillips 1997, p. 270).

Figure 7: Expected losses under alternative Data Generation Processes



Note: Expected losses generated by the different decision rules under alternative specifications for the mean (reported on the horizontal axis). For each mean, I generate 500 samples of 206 observations and replicate the same estimation as for the EuroStoxx50. The observations are drawn from the empirical distribution of the EuroStoxx50 time series. I then add different means to the sample, to simulate situations in which the judgmental decision of holding zero risky assets becomes less and less accurate. Expected losses are out-of-sample averages over the 500 samples for each mean.

Figure 8: Percentage of times statistical decisions underperform the judgmental decision under alternative Data Generation Processes



Note: Percentage of times the expected losses are greater than with the judgmental allocation, under alternative specifications for the mean (reported on the horizontal axis). The simulated data are the same as in Figure 5. Underperformance occurs more often when the judgmental allocation is close to the population mean. The maximum probability of underperformance is 100% for both Bayesian and maximum likelihood estimators.

the zero judgmental allocation.[9] When the population mean is equal to zero, subjective and pretest estimators underperform the same number of times. The underperformance rate does not coincide with the confidence levels of 1% and 10%, because for each simulated sample an out of sample exercise is conducted for the period January 2006 - December 2015. If one were to replicate this exercise only for one out of sample period, one would obtain an underperformance rate equal to the confidence level. As soon as one moves away from the zero mean, the underperformance rate of the pretest estimator deteriorates because it reverts to the maximum likelihood estimator. It is only for values of the population mean sufficiently far away from zero, that the underperformance rate starts to decline. The subjective classical estimator, instead, does not suffer from this drawback. Finally, the maximum likelihood and Bayesian estimators all have a maximum underperformance rate of 100%: when the judgmental allocation coincides with the population mean, allocations based on these estimators will underperform the judgmental allocation with probability one. In other words, these estimators cannot put an upper bound to the unconditional probability that their decision rule may be worse than the judgmental decision.

6 Conclusion

Bayesian statistics applies Bayes formula to combine a prior distribution with the likelihood distribution of the data, constructing a posterior distribution which exploits non sample and sample information. Bayesian decisions are obtained by minimizing the expected loss, using the posterior distribution to compute the expectation. In the decision space, this corresponds to a convex combination of the judgmental and maximum likelihood decisions, where the judgmental decision corresponds to the no data decision, that is the decision which minimizes the expected loss using the prior distribution. There must therefore exist a confidence interval around the maximum likelihood decision, whose edge coincides with the Bayesian decision. By making explicit the judgmental decision and the choice of the confidence level in a frequentist setting, it is possible to establish the equivalence between Bayesian and frequentist procedures.

[9]If one were to do this exercise conditional on the sample realization x (or equivalently on the observed *p-value* $\tilde{\alpha}$), the percentage of violations would of course correspond to $\alpha|x = g(\tilde{\alpha})$.

The confidence level is chosen as a mapping from the *p-value* of the first order conditions evaluated at the judgmental decision onto the unit interval. The frequentist decision maker selects decisions which are always at the boundary of the confidence interval. Beyond this boundary, the probability of obtaining higher expected losses than those implied by the judgmental decision becomes greater than the given confidence level. The confidence level reflects the attitude of the decision maker towards statistical uncertainty, referred to as statistical risk aversion. The incorporation of non sample information via prior distributions and Bayesian updating imposes time varying, decreasing statistical risk aversion on the decision maker, a fact which is at odds at least with my own preferences.

Appendix — Proofs

Proof of Theorem 2.1 — Consider only the case $i) - x + \tilde{a} \leq 0$. The other case can be proven in a similar way. If $\psi_1(x) = 0$, the null hypothesis $H_0 : -\theta + \tilde{a} \geq 0$ is not rejected at the given confidence level α. $\tilde{a}$ is therefore retained as the chosen action.

If $\psi_1(x) = 1$, the null hypothesis is rejected. Recalling its economic interpretation, rejection of the null implies that marginal moves away from $\tilde{a}$ by a sufficiently small amount $\Delta > 0$ decrease the loss function.

Consider now the family of null hypotheses of all the follow-up tests to $H_0 : -\theta + \tilde{a} \geq 0$, that is $H_\Delta : -\theta + \tilde{a} + \Delta \geq 0$ for $\Delta > 0$. Define also the family of rejection regions $R_\Delta \equiv \{\dot{x} \in \mathbb{R} : -\dot{x} + \tilde{a} + \Delta < c_{\alpha/2}\}$, where I have used the notation $\dot{x}$ to distinguish the potential realizations of the random variable X from the observed realization x. Clearly, $x \notin R_\Delta$ for any $\Delta \geq \bar{\Delta} \equiv c_{\alpha/2} + x - \tilde{a}$, that is the null hypothesis H_Δ is not rejected at the confidence level α for any $\Delta \geq \bar{\Delta}$.

Denote with $\hat{a}$ the chosen action and suppose that $\hat{a} \neq \tilde{a} + \bar{\Delta}$. If $\hat{a} = \tilde{a} + \Delta < \tilde{a} + \bar{\Delta}$, this implies that $x \in R_\Delta$, that is $H_\Delta : -\theta + \tilde{a} + \Delta \geq 0$ is rejected. Therefore this decision cannot be chosen.

If $\hat{a} = \tilde{a} + \Delta > \tilde{a} + \bar{\Delta}$, continuity implies that it exists $\epsilon > 0$ such that the null $H_{\Delta-\epsilon} : -\theta + (\hat{a} + \Delta - \epsilon) \geq 0$ was rejected at the given confidence level α, even though $x \notin R_{\Delta-\epsilon}$, which implies a contradiction.

The chosen action must therefore be $\hat{a} = \tilde{a} + \bar{\Delta} = c_{\alpha/2} + x$. $\square$

Proof of Theorem 2.2 — Consider the case when $-x + \tilde{a} \leq 0$. The other case is similar.

$$
\begin{aligned}
L(\theta, \delta^{A}(x)|\psi_1(x) = 0) - L(\theta,&\delta^{A}(x)|\psi_1(x) = 1) \\
&= -\theta(\tilde{a} - x - c_{\alpha/2}) + 0.5(\tilde{a}^2 - (x + c_{\alpha/2})^2)
\end{aligned}
$$

This function is linear in θ and therefore changes sign only once as a function of θ, specifically at the finite value $\theta = 0.5(\tilde{a} + x + c_{\alpha/2})$. Since $\psi_1(x)$ is a monotone procedure, the conditions of theorem 4 of Karlin and Rubin (1956) are satisfied and the result follows. $\square$

Proof of Theorem3.1 — Let's find out first the values of a for which $L(\theta, a) > L(\theta, \tilde{a})$. This is equivalent to finding out when the function $-a\theta + 0.5a^2 + \tilde{a}\theta - 0.5\tilde{a}^2$ is positive, which it is for $a < \theta - |-\theta + \tilde{a}|$ and $a > \theta + |-\theta + \tilde{a}|$. Therefore:

$$
\begin{aligned}
P_\theta(L(\theta, \delta^{A|x}(X)) > L(\theta, \tilde{a})) = P_\theta(&\delta^{A|x}(X) < \theta - |-\theta + \tilde{a}|) + &\quad (13)\\
&+ P_\theta(\delta^{A|x}(X) > \theta + |-\theta + \tilde{a}|) &\quad (14)
\end{aligned}
$$

Consider again only the case $i) -x + \tilde{a} \leq 0$, as the other one is similar, and note that in this case $\delta^{A|x}(X)$ can be rewritten as:

$$
\delta^{A|x}(X) = \tilde{a} + (x + c_{\alpha/2|x} - \tilde{a})\psi_1^A(X)
$$

Suppose first that $-\theta + \tilde{a} > 0$. Substituting the decision rule and rearranging the terms, probability (13) is equal to:

$$
\begin{aligned}
P_\theta(\delta^{A|x}(X) < \theta - |-\theta + \tilde{a}|) &= P_\theta(\delta^{A|x}(X) < 2\theta - \tilde{a}) \\
&= P_\theta((x + c_{\alpha/2|x} - \tilde{a})\psi_1^A(X) < 2\theta - 2\tilde{a}) \\
&= 0
\end{aligned}
$$

because $(x + c_{\alpha/2|x} - \tilde{a})\psi_1^A(X) \geq 0$, and probability (14) is equal to:

$$
\begin{aligned}
P_\theta(\delta^{A|x}(X) > \theta + |-\theta + \tilde{a}|) &= P_\theta(\delta^{A|x}(X) > \tilde{a}) \\
&= P_\theta((x + c_{\alpha/2|x} - \tilde{a})\psi_1^A(X) > 0) \\
&= P_\theta(\psi_1^A(X) = 1) \\
&\leq \alpha
\end{aligned}
$$

because by Bayes rule:

$$
\begin{aligned}
P_\theta(\psi_1^A(X)=1) &= \frac{P_\theta(-X+\tilde{a}<c_{\alpha/2|x})}{P_\theta(-X+\tilde{a}\le 0)} \\
&= \frac{P_\theta(-X+\theta<c_{\alpha/2|x}+\theta-\tilde{a})}{P_\theta(-X+\theta\le\theta-\tilde{a})} \\
&\le 2P_\theta(-X+\theta<c_{\alpha/2|x}) \\
&= \alpha
\end{aligned}
$$

where the inequality follows from the fact that the case currently analyzed is $-\theta+\tilde{a}>0$.

Suppose now that $-\theta+\tilde{a}<0$. Following a similar procedure, probability (13) is equal to:

$$
\begin{aligned}
P_\theta(\delta^{A|x}(X)<\theta-|-\theta+\tilde{a}|) &= P_\theta(\delta^{A|x}(X)<\tilde{a}) \\
&= P_\theta((x+c_{\alpha/2|x}-\tilde{a})\psi_1^A(X)<0) \\
&= 0
\end{aligned}
$$

and probability (14) is equal to:

$$
\begin{aligned}
P_\theta(\delta^{A|x}(X)>\theta+|-\theta+\tilde{a}|) &= P_\theta(\delta^{A|x}(X)>2\theta-\tilde{a}) \\
&= P_\theta((x+c_{\alpha/2|x}-\tilde{a})\psi_1^A(X)>2\theta-2\tilde{a}) \\
&\le P_\theta(\psi_1^A(X)=1) \\
&\le \alpha
\end{aligned}
$$

□

Proof of Proposition 4.1 — Consider the simplified decision rule (6). If $-x+\tilde{a}\le 0$, $\delta^{A|x}(x)=x+c_{\tilde{\alpha}}h=x+(-x+\tilde{a})h$. If $-x+\tilde{a}\ge 0$, $\delta^{A|x}(x)=x+c_{1-\tilde{\alpha}}h=x+(-x+\tilde{a})h$. □

Proof of Theorem 4.1 — Impose that (6) is equal to $\delta^\mu(x)$. This is possible because I have assumed that $\delta^\mu(x)$ shrinks from $\tilde{a}$ towards x. Solving for $\alpha|x$ gives the result. □

References

Al-Najjar, N. and J. Weinstein (2009), The Ambiguity Aversion Literature: A Critical Assessment, *Economics and Philosophy*, 25, 249-284.

Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis* (2nd ed.), New York: Springer-Verlag.

Bewley, T.F. (1988), Knightian Decision Theory and Econometric Inference, Yale University, Cowles Foundation Discussion Paper N. 868.

Bewley, T.F. (2011), Knightian Decision Theory and Econometric Inference, *Journal of Economic Theory*, 146 (3), 1134-1147.

Brandt, M.W. (2009), Portfolio Choice Problems, in Y. Ait-Sahalia and L. P. Hansen (eds.), *Handbook of Financial Econometrics*, North Holland.

Cerreia Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2013a), Ambiguity and Robust Statistics, *Journal of Economic Theory*, 148, 974-1049.

Cerreia Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2013b), Classical Subjective Expected Utility, *Proceedings of the National Academy of Sciences*, 110, 6754-6759.

Chamberlain, G., (2000), Econometrics and decision theory, *Journal of Econometrics*, 95 (2), 255-283.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009), Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy? *Review of Financial Studies*, 22, 1915-1953.

G. Elliott and A. Timmermann (2016), *Economic Forecasting*, Princeton University Press.

Ellsberg, D. (1961), Risk, Ambiguity, and the Savage Axioms, *The Quarterly Journal of Economics*, 75 (4), 643âĂŞ669.

Geweke, J. and C. Whiteman (2006), Bayesian Forecasting, in *Handbook of Economic Forecasting, Volume I*, edited by G. Elliott, C. W. J. Granger and A. Timmermann, Elsevier.

Gilboa, I. (2009), *Theory of Decision Under Uncertainty*, Cambridge University Press.

Gilboa, I. and M. Marinacci (2013), Ambiguity and the Bayesian Paradigm, in *Advances in Economics and Econometrics: Theory and Applications*, Tenth World Congress of the Econometric Society, D. Acemoglu, M. Arellano and E. Dekel (Eds.), New York, Cambridge University Press.

Gilboa, I. and D. Schmeidler (1989), Maxmin expected utility with non-unique prior, *Journal of Mathematical Economics*, 18 (2), 141-153.

Granger, C.W.J. and P. Newbold (1986), *Forecasting Economic Time Series*, Academic Press.

Guidolin, M. and F. Rinaldi (2013), Ambiguity in asset pricing and portfolio choice: a review of the literature, *Theory and Decision*, 74 (2), 183-217

Jobson, J.D. and B. Korkie (1981), Estimation for Markowitz Efficient Portfolios, *Journal of the American Statistical Association*, 75, 544-554.

Karlin, S. and H. Rubin (1956), The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio, *The Annals of Mathematical Statistics*, 27, 272-299.

Magnus, J.R. (2002), Estimation of the Mean of a Univariate Normal Distribution With Unknown Variance, *Econometrics Journal*, 5, 225-236.

Manganelli, S. (2009), Forecasting with Judgment, *Journal of Business and Economic Statistics*, 27 (4), 553-563.

Manski, C.F. (2000), Identification Problems and Decisions under Ambiguity: Empirical Analysis of Treatment Response and Normative Analysis of Treatment Choice, *Journal of Econometrics*, 95 (2), 415-442.

Manski, C.F. (2004), Statistical treatment rules for heterogeneous populations, *Econometrica*, 72 (4), 1221-1246.

Manski, C.F. (2013), *Public policy in an uncertain world: analysis and decisions*, Harvard University Press.

Dominitz, J., and C.F. Manski, C. (2017), More Data or Better Data? A Statistical Decision Problem, *The Review of Economic Studies*, 84 (4), 1583-1605.

Markowitz, H.M. (1952), Portfolio Selection, *Journal of Finance*, 39, 47-61.

Marinacci, M. (2015), Model Uncertainty, *Journal of European Economic Association*, 13, 998-1076.

Neyman, J. (1952), *Lectures and Conferences on Mathematical Statistics and Probability*, Graduate School U.S. Department of Agriculture.

Newey, W.K., and D. McFadden (1994), Large Sample Estimation and Hypothesis Testing, in *Handbook of Econometrics*, 4, 2111-2245.

Patton, A.J. and A. Timmermann (2007), Testing Forecast Optimality under Unknown Loss, *Journal of American Statistical Association*, 102, 1172-1184.

Patton, A.J. and A. Timmermann (2012), Forecast Rationality Tests Based on Multi-Horizon Bounds, *Journal of Business and Economic Statistics*, 30 (1), 1-17.

Phillips, P.C.P. (1997), The ET Interview: Professor Clive Granger, *Econometric Theory*, 13, 253-303.

Savage, L.J. (1954), *The Foundations of Statistics*, New York, John Wiley & Sons.

Schmeidler, D. (1989), Subjective probability and expected utility without additivity, *Econometrica*, 57: 571-587.

Siniscalchi, M. (2009), Two Out of Three Ain't Bad: A Comment on the 'Ambiguity Aversion Literature: A Critical Assessment', *Economics and Philosophy*, 25, 335-356.

Smets, F. and R. Wouters (2007), Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach, *American Economic Review*, 97 (3), 586-606.

Wald, A. (1950), *Statistical Decision Functions*, New York, John Wiley & Sons.

White, H. (1994), *Estimation, Inference and Specification Analysis*, Cambridge Books, Cambridge University Press.