# Deciding with Judgment

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

July, 2017

**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 (the decision taken in the absence of data) 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 also be interpreted as decisions under ambiguity.

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

---

*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, as well as seminar participants at the ECB, Humboldt University Berlin, Technische Universitaet Dresden, and the University of Bern.

# 1 Introduction

Non sample information is incorporated in statistical analysis in the form of prior probabilities. Bayes formula provides the optimal rule to combine non sample and sample information in the form of a posterior distribution. When applied to decision problems, the optimal decision minimizes the expected loss, using the posterior distribution to compute the expectation. The appeal of the Bayesian approach is that it can be justified on the basis of axiomatic foundations (Savage, 1954). A major drawback is that its application to statistical problems requires the decision maker to formulate her prior distribution, which is usually unknown and difficult to specify. In fact, lack of knowledge of the priors was one of the motivations to develop classical statistics.[1]

Frequentist statistics, on the other hand, has no explicit role for non sample information, not least because it was motivated by a desire for objectivity. The frequentist procedure, nevertheless, hides at its core non sample information: the choice of the hypothesis to be tested and the associated confidence level are ultimately driven by non sample considerations. This paper shows how the explicit treatment of non sample information provides a surprising bridge to connect frequentist and Bayesian statistics.

The starting point is the explicit formulation of non sample information as *judgment*, which is formed by a *judgmental decision* and the *confidence level* attached to it. It then turns the standard frequentist procedure on its head. Instead of testing whether the frequentist estimator is different from a pre-specified value, it starts from the judgmental decision and, for the given confidence level, tests whether the decision can be improved upon, according to a certain loss function. Careful consideration of the hypothesis to be tested reveals that the optimal decision is either the judgmental decision itself (in case the null hypothesis is not rejected) or it is at the boundary of the confidence interval.

The economic intuition, as also discussed in Manganelli (2009), is the following. The sample gradient of the loss function represents the empirical

[1]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).

approximation of the first order conditions of the loss minimization problem. It is a random variable and its realization will be different from zero, even when the judgmental decision coincides with the optimal decision. Testing whether the judgmental decision is optimal is equivalent to testing whether the sample gradient evaluated at the judgmental decision is statistically different from zero. If the judgmental decision is optimal, the probability that this statistic falls outside the confidence interval is equal to the chosen confidence level. In all the other cases, this probability is higher than that. Rejection of the null hypothesis provides statistical evidence that marginal moves away from the judgmental decision decrease the loss function in population. This holds until the boundary of the confidence interval is reached. Accordingly, the optimal frequentist decision prescribes to keep the judgmental decision if the realization of the test statistic falls within the confidence interval, and move to the closer edge of the confidence interval otherwise.

The key and novel element to establish the equivalence between Bayesian and frequentist statistics is the choice of the confidence level. I show how the most common estimators – maximum likelihood, pretest, Bayesian – can be interpreted as the optimal frequentist decision rule, when the confidence level is chosen as a function of the *p-value* of the first order conditions evaluated at the judgmental decision. For instance, a very low *p-value* may convince the decision maker to have little confidence in her initial judgment and induce her to adopt higher confidence levels. Pretest estimators are an extreme case in point. *P-values* higher than a given probability induce the decision maker to stick to her judgment. *P-values* lower than that probability (even infinitesimally lower) induce the decision maker to neglect her judgment and adopt instead the maximum likelihood estimator. The maximum likelihood decision always neglects any judgment by choosing a 0% confidence interval, which coincides with the decision setting the empirical gradient equal to zero.

To understand the link with Bayesian decision rules, consider that the confidence level determines the width of the confidence interval and therefore the amount of shrinkage towards the maximum likelihood decision, since the optimal frequentist decision always moves from the judgmental decision to the closest boundary of the confidence interval. The shrinkage factor is zero if the confidence level is equal to the *p-value* and it is one if the confidence interval is zero. 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. Since the *p-value* is a function of the sample realization, the frequentist choice of the confidence level is conditional on the observed sample. It is this conditioning that provides the fundamental link between Bayesian and frequentist statistics.

Frequentist confidence intervals have an intuitive economic interpretation, which can help the decision maker elicit the confidence level. From the perspective adopted in this paper, Type I errors are produced by statistical decision rules which perform worse than the judgmental decision. Since the confidence level controls the probability of Type I errors, decisions which move beyond the edge of the confidence interval will underperform the judgmental decision with probability greater than the chosen confidence level. This suggests an alternative interpretation of the confidence level, as the willingness of the decision maker to take statistical risks (i.e. to commit Type I errors), which I refer to as *Statistical Risk Propensity*. Notice that this concept is distinct from the standard concept of risk aversion characterizing utility functions. If the Bayesian approach provides the theoretical foundations of statistics, the frequentist interpretation defines its purpose: help decision makers to take better decisions, by providing statistical decision rules which will not underperform the judgmental decision with a given probability.

I have argued so far that for any given prior there is a choice of judgment which makes Bayesian and frequentist decisions equivalent. The converse is not true. For any given judgment — which, recall, it is formed by a judgmental decision and a confidence level — there is a whole set of priors which are consistent with it. These are all the priors which give: 1) the judgmental decision in case of a decision problem without data and 2) for a given sample realization, all the frequentist decisions associated with the infinite set of confidence intervals with coverage probability equal to one minus the confidence level contained in the judgment. I show that the optimal frequentist decision is equivalent to minimizing the maximum expected loss over this set of priors. Frequentist decisions can therefore be interpreted as ambiguity averse decisions, in the sense of Gilboa and Schmeidler (1989).

I use an asset allocation problem as a working example to illustrate the 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. Several decision rules are implemented and compared, using an out of sample exercise. The results confirm a well known fact in the empirical asset allocation literature, that it is difficult to find allocations with good out of sample performance. The weight associated with the maximum likelihood decision rule is the most volatile. By the end of the sample, an investor following this investment strategy would have lost about one quarter of her initial wealth. The Bayesian decision rules perform slightly better, as the weights are shrunk towards zero, but would still have lost between 9% and 12%. The best performing decision rules are those that recommend to stick with the initial judgment of holding only cash, because the data is just too noisy to suggest a significant departure from it.

The insight of this exercise, however, is not to claim the superiority of some decision rules relative to others: like investors with different risk aversion in their utility function, investors with different confidence levels (or equivalently, with different priors) choose different allocations. A Monte Carlo exercise shows that decision rules characterized by lower confidence levels (and therefore larger confidence intervals) perform better when the initial judgment is close to the optimal one, but perform worse otherwise. 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.[2] The same Monte Carlo exercise reveals also that for both Bayesian and maximum likelihood decision rules the unconditional upper bound probability of performing worse than the judgmental allocation is 100%. In general, decision rules based on these estimators will perform worse than the judgmental allocation, when the judgmental allocation is very accurate.

The paper is structured as follows. The next section introduces judgment in frequentist decision theory. Section 3 establishes the equivalence between Bayesian and frequentist decision rules. Section 4 establishes the link between frequentist and ambiguity averse decisions. Section 5 presents the empirical evidence and section 6 concludes.

[2]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). A similar statement is reported by Geweke and Whiteman (2006) as opening quote of their paper, taken from Granger (1986).

## 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. It shows how hypothesis testing can be used to arrive at optimal frequentist decisions and shows how different choices of the confidence level associated with the hypothesis testing give rise to the most common estimators in econometrics.

For concreteness, I consider an asset allocation problem of an investor who is holding all her wealth in cash and has to decide what fraction $a \in \Re$ of her wealth to invest in a stock market index. Let $X_t \sim N(\theta, 1)$ be the distribution of the stock market index return, with known variance, but unknown expected return equal to the parameter $\theta \in \Re$. Let also the return on cash be zero. Suppose the investor wants to minimize a mean-variance loss function à la Markowitz (1952). The decision problem can be formalized as follows:

**Definition 2.1** (**Decision Problem**). *Let $X_t \sim N(\theta, 1)$, for $\theta \in \Re$, $t = 1, 2, ...$ Suppose that one realization of the random variable is available, $X_1 = x_1$. A decision maker would like to choose the action $a \in \Re$ such that:*

$$\min_a L(\theta, a) \tag{1}$$

*where $L(\theta, a) \equiv -a\theta + 0.5a^2$.*

The loss function $L(\theta, a)$ summarizes the *physical* uncertainty facing the decision maker, using the terminology of Marinacci (2015). It can be interpreted as the expectation, for a given value of parameter $\theta$, of the negative of a von Neuman-Morgenstern utility function with respect to the random variable $X_2$: $-u(X_2, a) = -aX_2 + 0.5a^2$. Formally, if $\Phi(x|\theta)$ denotes the cdf of $X_2$ conditional on $\theta$, then $L(\theta, a) = \int [-ax + 0.5a^2] d\Phi(x|\theta)$.

The problem needs more structure to be solved, as the parameter $\theta$ is unknown. The additional structure comes in the form of subjective information of the decision maker about $\theta$, which is referred to by Marinacci (2015) as *epistemic* uncertainty, from the Greek word for knowledge.

Before introducing additional structure, consider the following standard definition of a decision rule (Berger, 1985).

**Definition 2.2** (**Decision Rule**). *$\delta(X_1) : \Re \to \Re$ is a decision rule, such that if $X_1 = x_1$ is the sample realization, $\delta(x_1)$ is the action that will be taken.*

## 2.1 The Bayesian Decision

The Bayesian solution to the Decision Problem 2.1 assumes that the decision maker uses non sample information in the form of a prior distribution over the unknown parameter $\theta$:

**Definition 2.3** (**Prior**). *The non sample information of the decision maker is summarized by the prior cdf $\mu(\theta)$ over the parameter $\theta \in \Re$.*

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

**Proposition 2.1** (**Bayesian Decision**). *Assume the decision maker knows her prior distribution $\mu(\theta)$. The Bayesian solution to the Decision Problem 2.1 is:*

$$\delta^{\mu}(x_1) = \arg\min_{a} \int L(\theta, a) d\mu(\theta|x_1) \tag{2}$$

*where $\mu(\theta|x_1)$ denotes the posterior distribution.*

**Proof** — It follows immediately from the subjective expected utility theorem and application of Bayes rule. □

The Bayesian decision has the considerable merit of having an axiomatic justification, being grounded on a decision theoretic framework.[3] The main criticism to the Bayesian approach is that such a prior distribution is usually unknown and very difficult to specify. Without the specification of a prior, the Bayesian procedure is not applicable. Vague knowledge of the prior is not enough, as it is well known that slightly different priors may generate substantially different decisions.

## 2.2 The Frequentist Decision

Classical statistics as developed by Neyman and Fisher has no explicit role for *epistemic* uncertainty, as it was motivated by the desire for objectivity. Non sample information is, nevertheless, implicitly introduced in various forms,

[3]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.

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, which will be discussed in the next section.

#### 2.2.1 Judgment

In a frequentist setting, there is no prior distribution to help solving the Decision Problem 2.1. One solution often used is the plug-in estimator, which replaces the unknown $\theta$ parameter with its sample counterpart (see, for instance, chapter 3 of Elliott and Timmermann, 2016). In the context of the Decision Problem 2.1, this is given by $x_1$. Once $\theta$ is replaced by $x_1$, the optimal decision is $\delta^{Plug-in}(x_1) = x_1$. The problem with such decision is that it does not minimize the loss function, but rather its sample equivalent, and neglects estimation error.

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

**Definition 2.4** (**Judgment**)**.** *The non sample information of the decision maker is summarized by the pair $A \equiv (\tilde{a}, \alpha)$, which is referred to as the **judgment** of the decision maker. $\tilde{a} \in \Re$ is the **judgmental decision**, that is the action that the decision maker would take before seeing the data. $\alpha \in [0, 1]$ is the **confidence level** of the decision maker in the judgmental decision.*

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%). More generally, one can think of $\tilde{a}$ as the action that decision makers without access to statistical analysis would choose. It is essential for the development of the theory that $\tilde{a}$ is not random, as otherwise the statistical procedure that follows would be invalid. 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 decision $\tilde{a}$ which is observationally equivalent to a Bayesian decision with no data: $\tilde{a}$ solves the problem $\min_a \int L(\theta, a) d\mu(\theta)$.

$\alpha$ 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 $\alpha$ is also closely linked to the choice of prior distributions and determines the updating mechanism in a frequentist setting, after observing the realization $X_1 = x_1$. The decision in the light of the statistical evidence $x_1$ is taken on the basis of hypothesis testing.

#### 2.2.2 Hypothesis Testing

Consider a frequentist decision maker with judgment $A \equiv (\tilde{a}, \alpha)$ facing the Decision Problem 2.1. Replacing $\theta$ with its sample counterpart and considering the first order conditions evaluated at $\tilde{a}$, one obtains the following test statistic: $-X_1 + \tilde{a}$, which is distributed as a standard normal distribution, under the null hypothesis that $\tilde{a}$ is optimal. Given the confidence level $\alpha$ included in the judgment, it is possible for the frequentist decision maker to test whether $\tilde{a}$ is optimal. If the null hypothesis $H_0 : -\theta + \tilde{a} = 0$ cannot be rejected, then $\tilde{a}$ is the optimal decision, given the judgment and the sample realization. Rejection of the null hypothesis, on the other hand, has the economic interpretation that $\tilde{a}$ does not satisfy the first order conditions and therefore that marginal moves away from $\tilde{a}$ are likely to decrease the loss function. This holds until the boundary of the confidence interval is reached. The following proposition formalizes this reasoning.

**Proposition 2.2 (Hypothesis Testing).** *Assume the decision maker has judgment $A \equiv (\tilde{a}, \alpha)$. The frequentist solution to the Decision Problem 2.1 is:*

$$\begin{aligned}\delta^A(x_1) = {} & (x_1 + c_{\alpha/2}) I(-x_1 + \tilde{a} \leq c_{\alpha/2}) + \\ & + \tilde{a} I(c_{\alpha/2} < -x_1 + \tilde{a} < c_{1-\alpha/2}) + \\ & + (x_1 + c_{1-\alpha/2}) I(-x_1 + \tilde{a} \geq c_{1-\alpha/2})\end{aligned} \tag{3}$$

*where $c_\alpha = \Phi^{-1}(\alpha)$ and $\Phi(\cdot)$ is the cdf of the standard Normal distribution.*

**Proof** — See Appendix.

To understand the intuition behind this decision rule, suppose that $x_1$ is less than $\tilde{a}$, so that $-x_1 + \tilde{a} > 0$. This implies that lower values of $\tilde{a}$, i.e.

moves towards the maximum likelihood estimator $x_1$, decrease the sample analogue of the loss function. The decision maker would like to rule out that the opposite is true in population, to avoid that deviating from her initial judgment would make her worse off. The null and alternative hypotheses to be tested when $-x_1 + \tilde{a} > 0$ are:

$$H_0 : -\theta + \tilde{a} \leq 0 \qquad\qquad H_1 : -\theta + \tilde{a} > 0 \tag{4}$$

Frequentist hypothesis testing is based on the following thought experiment. Suppose data are generated by values of $\theta$ consistent with the null hypothesis, what is the probability of obtaining the observed sample realization, $-x_1 + \tilde{a}$? If this probability is lower than the chosen (and subjective) confidence level $\alpha$, $H_0$ is rejected and the decision maker can be confident that by moving from $\tilde{a}$ towards $x_1$ she decreases her loss function. This reasoning holds until the boundary of the confidence interval is reached, which gives the estimator incorporating judgment.

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 $\alpha$, 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 $\tilde{a}$ to $x_1$, 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 $1 - \beta(\theta)$, where $\beta(\theta)$ 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 $\alpha$ generally implies also a small power $\beta(\theta)$ for values of $\tilde{a}$ close to $\theta$. 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.

One final remark is that the set up of this example can be easily generalized to incorporate standard results from asymptotic theory. Suppose that a sample $\{x_t\}_{t=1}^{T}$ is observed and that a maximum likelihood estimator is computed, $\hat{\theta}_T$. A mean value expansion of the gradient of the loss function around the true parameter $\theta$ yields:

$$\nabla_a L(\hat{\theta}_T, \tilde{a}) = \nabla_a L(\theta, \tilde{a}) + \nabla_{a,\theta} L(\bar{\theta}, \tilde{a})(\hat{\theta}_T - \theta) \tag{5}$$

Table 1: Hypothesis testing

| | | **Decision** $H_0$ | $H_1$ |
|---|---|---|---|
| **Truth** | $H_0$ | Avoid higher loss<br>$1-\alpha$ | Higher loss<br>$\alpha$ |
| | $H_1$ | Fail to lower loss<br>$1-\beta(\theta)$ | Lower loss<br>$\beta(\theta)$ |

*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 the population gradient has the same sign as the sample gradient.
$\alpha$ and $\beta(\theta)$ are the size and power of the test.

where $\bar{\theta}$ is a mean value which lies between $\theta$ and $\hat{\theta}_T$. Under standard regularity conditions (see for instance Newey and McFadden, 1994, or White, 1994), $\sqrt{T}(\hat{\theta}_T - \theta) \sim N(0, \Sigma)$. Under the null hypothesis that $\tilde{a}$ is the optimal decision $\nabla_a L(\theta, \tilde{a}) = 0$ and therefore:

$$\sqrt{T}\nabla_a L(\hat{\theta}_T, \tilde{a}) \sim N(0, \nabla_{a,\theta} L(\theta, \tilde{a})\Sigma\nabla'_{a,\theta} L(\theta, \tilde{a})) \tag{6}$$

The hypothesis testing procedure described above can now be applied to the normally distributed random variable $\sqrt{T}\nabla_a L(\hat{\theta}_T, \tilde{a})$.

#### 2.2.3 Choosing the Confidence Level

Careful choice of the confidence level $\alpha$ of the judgment pair $A \equiv (\tilde{a}, \alpha)$ generates the most common estimators in econometrics as a special case of the decision rule $\delta^A(x_1)$. A key insight of this subsection is that although $\alpha$ 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 $x_1$.

The choice of the confidence level $\alpha$ may generally be considered as a mapping into the interval $[0,1]$ from the *p-value* of the test statistic under the hypothesis that $\tilde{a}$ is the optimal decision.

**Definition 2.5** (**Choice of the Confidence Level**). *Define* $\tilde{\alpha} \equiv \Phi(-x_1 + \tilde{a})$*, the* p-value *of the test statistic* $\nabla_a L(X_1, \tilde{a})$ *associated with the judgmental decision* $\tilde{a}$*, under the null hypothesis that* $\nabla_a L(\theta, \tilde{a}) = 0$*. The decision maker chooses the confidence level as follows:*

$$\alpha|x_1 = g(\tilde{\alpha}) : [0,1] \to [0,1] \tag{7}$$

Since the *p-value* $\tilde{\alpha}$ is determined by the sample realization $x_1$, the choice of $\alpha$ is conditional on $x_1$. I have made explicit this fact with the notation $\alpha|x_1$.

Any judgmental decision $\tilde{a}$ with *p-value* $\tilde{\alpha} \geq \alpha$ if $\tilde{\alpha} \leq 0.5$ (or $\tilde{\alpha} \leq \alpha$ if $\tilde{\alpha} \geq 0.5$) implies that the decision taken is the judgmental decision itself, because in this case $(-x_1 + \tilde{a})$ falls within the confidence interval of the decision rule (3). One can therefore impose the condition that $g(\tilde{\alpha}) \geq \tilde{\alpha}$ if $\tilde{\alpha} \leq 0.5$ and $g(\tilde{\alpha}) \leq \tilde{\alpha}$ if $\tilde{\alpha} \geq 0.5$ without affecting the optimal decision. This condition ensures that decisions are always at the boundary of the confidence interval. Note that in this case the decision rule (3) simplifies to:

$$\delta^{A|x_1}(x_1) = x_1 + c_{\alpha|x_1} \tag{8}$$

where the notation $\delta^{A|x_1}$ is equivalent to $\delta^A$, but makes explicit the conditioning on $x_1$. The important thing to notice is that this decision rule depends on $x_1$ in two ways: directly, as it is a function of the sample realization, and indirectly, via the choice of the confidence level $\alpha$. When computing the risk function of $\delta^{A|x_1}(X_1)$, the argument $X_1$ is random, but the confidence level will be conditioned on the observed sample realization $x_1$. It is this conditioning that provides the fundamental link between Bayesian and frequentist statistics.

Here are some common examples of how the function $g(\tilde{\alpha})$ in (7) is chosen:

1. <u>Maximum likelihood</u>:

$$\alpha|x_1 = 0.5, \forall\ \tilde{\alpha} \in [0,1]$$

2. <u>Pretest estimator with threshold $\bar{\alpha}$</u>:

$$\alpha|x_1 = \begin{cases} 0.5 & \text{if } \tilde{\alpha} \in [0, \bar{\alpha}/2) \cup (1-\bar{\alpha}/2, 1] \\ \tilde{\alpha} & \text{if } \tilde{\alpha} \in [\bar{\alpha}/2, 1-\bar{\alpha}/2] \end{cases}$$

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

$$\alpha|x_1 = \begin{cases} \bar{\alpha}/2 & \text{if } \tilde{\alpha} \in [0, \bar{\alpha}/2) \\ 1-\bar{\alpha}/2 & \text{if } \tilde{\alpha} \in (1-\bar{\alpha}/2, 1] \\ \tilde{\alpha} & \text{if } \tilde{\alpha} \in [\bar{\alpha}/2, 1-\bar{\alpha}/2] \end{cases}$$

4. Judgmental decision:

$$\alpha|x_1 = \tilde{\alpha}, \forall\ \tilde{\alpha} \in [0,1]$$

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 (8). These examples clarify how the framework encompasses the most common estimators. The maximum likelihood estimator always disregards any judgmental decision, by setting the confidence level equal to 0.5. In this case, $c_{\alpha|x_1} = 0$ and $\delta^{A|x_1}(x_1) = x_1$. 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 0.5 otherwise. The subjective classical estimator maintains the threshold probability $\bar{\alpha}$ for *p-values* in the tails of the interval, 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 $\tilde{\alpha}$.

In general, the choice of the confidence level $\alpha|x_1$ can be any function of the *p-value* of the test statistic. This is the case for Bayesian estimators, as will be shown next.

# 3 Equivalence between Bayesian and Frequentist Decisions

Careful choice of the confidence level $\alpha|x_1$ allows one to arrive at frequentist decisions which are equivalent to Bayesian decisions.

Let's start by noticing that the decision rule $\delta^{A|x_1}(x_1)$ of equation (8) is a shrinkage estimate.

**Proposition 3.1** (**Skrinkage**). *Given the judgment* $A|x_1 \equiv (\tilde{a}, \alpha|x_1)$*, for any sample realization* $x_1$*, the decision rule* $\delta^{A|x_1}(x_1)$ *of Proposition 2.2 is a shrinkage estimate of the type* $\delta^{A|x_1}(x_1) = (1-h)x_1 + h\tilde{a}$*, where* $h \equiv c_{\alpha|x_1}/c_{\tilde{\alpha}} \in [0,1]$.

**Proof** — See Appendix.

This decision rule is a convex combination between the judgmental decision and the maximum likelihood estimator. The amount of shrinkage is determined by the factor $h$, which is a combination of data (as represented by $x_1$) and judgmental information (as represented by the judgmental decision $\tilde{a}$ and the associated confidence level $\alpha|x_1$). 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 $X_t \sim N(\theta, 1)$, the posterior mean is $[\tau^2/(1+\tau^2)]x_1 + [1/(1+\tau^2)]\tilde{a}$. Informative priors (that is, low $\tau^2$) imply a posterior mean close to the prior mean. In Proposition 3.1, judgment with low confidence level (that is, an $\alpha|x_1$ 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 $\tau^2$) imply a posterior mean close to the maximum likelihood estimator, which in turn is equivalent to judgment with high confidence level (that is, an $\alpha|x_1$ close to 0.5 and $h$ close to 0).

The following proposition shows that for a Bayesian decision associated with a given prior there is a corresponding choice of $\alpha|x_1 = g(\tilde{\alpha})$ which produces an equivalent frequentist decision.

**Proposition 3.2 (Equivalence between Bayesian and Frequentist Decisions).** *For a given prior distribution $\mu(\theta)$ such that $\tilde{a} = \arg\min_a \int L(\theta, a)\, d\mu(\theta)$ and $\delta^\mu(x_1)$ lies between $\tilde{a}$ and $x_1$, the Bayesian solution to the Decision Problem 2.1 is equivalent to the frequentist solution $\delta^{A|x_1}(x_1)$, when*

$$\alpha|x_1 = \Phi(\delta^\mu(x_1) - x_1) \tag{9}$$

*where $x_1 \equiv \tilde{a} - \Phi^{-1}(\tilde{\alpha})$.*

**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 $\theta$ is Normally distributed with mean zero and variance $1/c$, the optimal Bayesian decision is:

$$\delta^N(x_1) = (1+c)^{-1}x_1 \tag{10}$$

**Bayesian estimator based on Laplace prior -** If the prior over the parameter $\theta$ is distributed as a Laplace with mean zero and scale parameter $c$, the optimal Bayesian decision is:

$$\delta^L(x_1) = x_1 - c \cdot \frac{1 - \exp(2cx_1)\frac{\Phi(-x_1-c)}{\Phi(x_1-c)}}{1 + \exp(2cx_1)\frac{\Phi(-x_1-c)}{\Phi(x_1-c)}} \tag{11}$$

In figure 1, I compare the confidence levels associated with the decision rules discussed in section 2.2.3 and the two Bayesian decision rules above. See also Magnus (2002) and Manganelli (2009) for a detailed risk analysis of similar estimators.

Note how all statistical decision rules have a confidence level mapping which falls in between the two extreme situations: the judgmental decision (where no data is taken into consideration) and the maximum likelihood estimator (where no judgment is taken into consideration). The judgmental decision is described by the diagonal line in the space $(\alpha|x_1, \tilde{\alpha})$. As already discussed in section 2.2.3, any point below this diagonal line for $\tilde{\alpha} \leq 0.5$ (above the diagonal for $\tilde{\alpha} \geq 0.5$) is equivalent to its vertical projection on the diagonal, in the sense that they all imply that the optimal decision coincides with the judgmental decision $\tilde{a}$. The confidence level $\alpha|x_1$ associated with the maximum likelihood estimator, instead, does not depend on $\tilde{\alpha}$ and is always equal to 0.5.

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 5% and greater than 95% in the example of figure 1). 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. It is equivalent to the Burr estimator and, being kinked, it is not admissible (Magnus, 2002). I do not consider it a drawback of this decision rule, because, as argued often by Bayesians, the concept of admissibility fails to condition on the observed sample realization $x_1$. Nevertheless, I notice here that according to the complete class theorem, under some regularity conditions, every admissible rule is a generalized Bayes rule with respect to some prior $\mu(\theta)$. There must therefore be a connection between admissibility and the properties of the $g(\tilde{\alpha})$ function determining the confidence level of the frequentist decision maker.

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. As a decision maker I would personally behave in exactly the opposite way: when data prove my initial judgment to be extremely bad, I would revert to the maximum likelihood estimator and assign zero weight to my judgment.

Second, the two estimators are characterized by different confidence levels, despite being based on prior distributions which have been calibrated to have both zero mean and unit variance. As already evidenced 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. Berger (1985), for instance, raises similar issues by comparing decisions based on Normal and Cauchy priors matched to have the same median and interquartiles (see example 2, p. 111).

Third, 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.

Fourth, this paper shows that imposing priors on parameters is equivalent to imposing specific confidence levels to 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 2 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 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 accuracy of the estimator increases and therefore 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 a challenge for the updating of the confidence level mapping in a frequentist setting for dynamic problems: the optimal updating rule, when new data become available, relies on application of Bayes formula, which cannot be applied without knowledge of the prior distribution. I see two answers to this challenge. The first one is to use the confidence level mapping to elicit priors. First, the decision maker, for given non sample information and sample size, selects her confidence level function $g(\tilde{\alpha})$, and in a second stage the econometrician searches within a wide set of probability distributions the prior with the closest mapping with the decision maker's confidence level. The selected prior distribution can then be used to update the posterior and the confidence level mapping when new data arrive. The second answer is to acknowledge that additional data come with even greater non sample information. At each new time period, the decision maker may reasonably have different judgment (or, equivalently, different priors), so that each decision can be interpreted as a new static decision.[4]

# 4 Frequentist Decisions as Ambiguity Averse Decisions

I have discussed in the previous section that it is very challenging to elicit priors for a decision maker. In this section I argue that it is relatively easier to elicit judgment of the type $A|x_1 \equiv (\tilde{a}, \alpha|x_1)$. In most practical situations, the decision maker knows $\tilde{a}$, as it is the decision taken in the absence of statistics. This section shows that the confidence interval associated with the frequentist decision rule has an intuitive economic interpretation, which can help the decision maker to choose $\alpha|x_1$. The section also shows that the frequentist decision can be interpreted as a decision under ambiguity, where the decision maker chooses the best decision from the worst prior within a set of probability distributions compatible with the judgment $A|x_1$.

## 4.1 Confidence Intervals as the Purpose of Statistics

Let's assume that the decision maker does not know the precise form of $\mu(\theta)$, but that she is able to formulate her judgment as $A|x_1 \equiv \{\tilde{a}, \alpha|x_1\}$. I show that the confidence level $\alpha|x_1$ has a very intuitive economic interpretation, and should therefore be relatively easy to elicit.

Let's consider again the decision rule (3), so that $\alpha|x_1$ is the confidence level associated with a two-sided test. For a given $\alpha|x_1$, there is an infinity of confidence intervals with coverage probability $1 - \alpha|x_1$. Let's therefore

[4] 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 for non Bayesian approaches (see, for instance, Al-Najjar and Weinstein, 2009, and the follow up comment by Siniscalchi, 2009).

redefine the decision rule (3) in terms of these infinite choices:

$$\begin{aligned}\delta^{A|x_1}(\bar{\alpha}; x_1) \equiv\ & (x_1 + c_{\underline{\alpha}})I(-x_1 + \tilde{a} \leq c_{\underline{\alpha}}) + \\ & + \tilde{a}I(c_{\underline{\alpha}} < -x_1 + \tilde{a} < c_{1-\bar{\alpha}}) + \\ & + (x_1 + c_{1-\bar{\alpha}})I(-x_1 + \tilde{a} \geq c_{1-\bar{\alpha}})\end{aligned} \tag{12}$$

where $\underline{\alpha} = \alpha|x_1 - \bar{\alpha}$, $\bar{\alpha} \geq 0$ and $\underline{\alpha} \geq 0$. For any given $\bar{\alpha}$, this decision rule is one of the many $(1 - \alpha|x_1)$ confidence intervals associated with the random variable $-X_1 + \tilde{a}$ under the null hypothesis that $\tilde{a}$ is the optimal decision.

**Proposition 4.1** (**Confidence Intervals**). *Consider the Decision Problem 2.1 and assume the decision maker has judgment* $A|x_1 \equiv (\tilde{a}, \alpha|x_1)$. *The decision rule* $\delta^{A|x_1}(\bar{\alpha}; X_1)$ *performs worse than the judgmental decision* $\tilde{a}$ *with probability not greater than* $\alpha|x_1$*:*

$$P_\theta[L(\theta, \delta^{A|x_1}(\bar{\alpha}; X_1)) > L(\theta, \tilde{a})] \leq \alpha|x_1 \tag{13}$$

**Proof** — See Appendix.

Proposition 4.1 gives a frequentist interpretation to Bayesian decisions. Recall from Proposition 3.2 that for any given prior and sample realization $x_1$, there is a choice of $\alpha|x_1$ which makes Bayesian and frequentist decisions equivalent. For a given $\alpha|x_1$, one can therefore perform the following frequentist thought experiment: draw an infinite amount of samples $\{x_1^i\}_{i=1}^{\infty}$ from $X_1 \sim N(\theta, 1)$ and compute how many times the statistical decision rule $\delta^{A|x_1}(\bar{\alpha}; x_1^i)$ performs worse than the judgmental decision. This probability will depend on the unknown population value of $\theta$, but the proposition 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 uncertainty, the decision maker cannot be sure that this will be the case and the guarantee can only be expressed in terms of a probability.

Proposition 4.1 suggests an alternative interpretation of the confidence level. Since $\alpha|x_1$ 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 Problem 2.1. One could therefore alternatively refer to the confidence level $\alpha|x_1$ as to the degree of *Statistical Risk Propensity*.

If the Bayesian approach provides the theoretical foundations of statistics, the frequentist interpretation defines its purpose: help decision makers to take better decisions, by providing statistical decision rules which will not do worse than the judgmental decision with a given probability.

## 4.2 Ambiguity Aversion and Minimax Rules

The result of Proposition 4.1 holds for any arbitrary confidence intervals $[c_{\underline{\alpha}}, c_{1-\bar{\alpha}}]$, as long as they have coverage probability equal to $(1 - \alpha|x_1)$. The choice of $\underline{\alpha}$ and $\bar{\alpha}$ may in principle be left as a choice of the decision maker, but this subsection shows that the symmetric confidence interval has minimax properties, providing an argument for selecting this one out of the infinite possibilities.

Consider the standard definition of risk (see Berger 1985):

**Definition 4.1** (**Risk**). *The risk function of the decision rule $\delta^{A|x_1}(\bar{\alpha}; X_1)$ is defined by*

$$R(\theta, \bar{\alpha}) = E_\theta[L(\theta, \delta^{A|x_1}(\bar{\alpha}; X_1))] = \int L(\theta, \delta^{A|x_1}(\bar{\alpha}; x))d\Phi(x|\theta) \tag{14}$$

*where $\Phi(\cdot)$ is the cdf of $X_1$.*

The following proposition determines the minimax confidence interval.

**Proposition 4.2** (**Minimax Confidence Interval**). *Under the same conditions as Proposition 4.1, $\bar{\alpha} = \alpha|x_1/2$ is the solution to the following problem:*

$$\min_{\bar{\alpha} \in [0, \alpha|x_1]} \max_{\theta \in \Re} R(\theta, \bar{\alpha}) \tag{15}$$

*giving the decision rule $\delta^{A|x_1}(\alpha|x_1/2; x_1)$ in* (12)*.*

**Proof** — See Appendix.

The criterion is constrained minimax, as it considers the best possible action in the case the unknown parameter $\theta$ assumes the least favourable value, subject to the constraint that the statistical decision rule does not perform worse than the judgmental decision with probability $\alpha|x_1$. It represents an additional guarantee to the decision maker who decides to abandon her judgmental decision to follow a statistical procedure: it gives an upper bound to how worse the statistical procedure can be relative to the judgmental decision, which in the case of Proposition 4.2 is given by $R(\tilde{a}, \alpha|x_1/2)$. The symmetry of the optimal confidence interval depends on the specific loss function and on the fact that the distribution of $-X_1 + \theta$ is gaussian and therefore symmetric. If one were to use an asymmetric distribution, the optimal confidence interval would also be asymmetric.

The problem is related to the robust control problem of Hansen and Sargent (2008).[5] They are concerned with decision making in the presence of model misspecification. The decision maker bases her decision on a misspecified model, but is aware that her model is misspecified, and seeks a decision rule that is robust across a set of models within a pre-specified distance (measured by the relative entropy) from the approximating model. Robustness is achieved by considering the worst case shocks from the set of models, subject to the constraint that the relative entropy is not larger than a given value. The result of Proposition 4.2, instead, obtains robustness by considering the worst possible value for $\theta$, subject to the constraint that the chosen decision rule does not perform worse than the judgmental decision $\tilde{a}$ with a pre-specified probability $\alpha|x_1$.

The decision associated with the minimax confidence interval can be given the interpretation of an ambiguity averse decision, as per the axiomatic development of Gilboa and Schmeidler (1989). Consider the following set of priors compatible with the judgment $A|x_1$:

$$\Gamma(A|x_1) \equiv \{\mu(\theta) : \delta^{\mu} = \tilde{a} \text{ and } \delta^{\mu}(x_1) = \delta^{A|x_1}(\bar{\alpha}; x_1),\ \bar{\alpha} \leq \alpha|x_1\} \tag{16}$$

where $\delta^{\mu}$ denotes the Bayesian decision without data. $\Gamma(A|x_1)$ denotes the set of all the priors which give the decision $\tilde{a}$ in case of no data and the frequentist decision after observing the sample realization $x_1$. The following proposition shows that the frequentist decision associated with the minimax confidence interval is equivalent to an ambiguity averse decision.

**Proposition 4.3** (**Ambiguity Aversion**). *The optimal decision rule* $\delta^{A|x_1}$

[5] See the constraint problem stated in equation 2.4.1, p. 32.

$(\alpha|x_1/2; x_1)$ *of Proposition 4.2 is the solution of the following problem:*

$$\min_a \max_{\mu(\theta)\in\Gamma(A|x_1)} \int L(\theta, a) d\mu(\theta|x_1) \qquad (17)$$

**Proof** — See Appendix.

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.

# 5 Empirical evidence

Section 3 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 classical 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 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?

[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).

The interpretation of the confidence interval given in section 3 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 2 reports summary statistics.

Table 2: 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 estimators 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 $\{\tilde{x}_t\}_{t=1}^{T}$ the original time series of log returns, let $\sigma$ be the full sample standard deviation and $T_1 < T$ the size of the first estimation sample. Then, for each $T_1 + s$, $s = 0, 1, 2, ..., T - T_1 - 1$, define:

$$\{x_t\}_{t=1}^{T_1+s} = \{\sqrt{(T_1+s)}\tilde{x}_t/\sigma\}_{t=1}^{T_1+s} \quad \text{and} \quad \bar{x}_{T_1+s} = (T_1+s)^{-1}\sum_{t=1}^{T_1+s} x_t \qquad (18)$$

I 'help' the estimators 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 $\bar{x}_{T_1+s}$ is normally distributed with variance equal to one and unknown mean. We can therefore implement the estimators discussed in the preceding sections of the paper, using the single observation $\bar{x}_{T_1+s}$ for each period $T_1 + s$.

The results of this exercise are reported in figures 3 and 4. Figure 3 plots the optimal weights obtained from the different estimators. A few things are worth noticing. First, the weight associated with the maximum likelihood estimator is the most volatile, as it is the one that suffers the most from estimation error. The Bayesian estimators 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 1. Pretest and subjective classical estimators 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 estimator, which for that month coincides with the maximum likelihood estimator (remember that when the *p-value* is less than $\bar{\alpha}$, the pretest estimator reverts to the maximum likelihood estimator). The weight associated with the subjective classical estimator with 40% confidence threshold exhibits just a small blip, as it goes to the boundary of the confidence interval associated with $\alpha$ instead of moving all the way to the maximum likelihood.

Figure 4 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 estimator with threshold of 40% would have lost little less than 5%. Note that the entire loss comes from shorting the position and following the predictions of the maximum likelihood estimator in February 2009. In all the other months there is no investment in the stock market. The other three estimators – the pretest with 1% and the subjective classical estimators 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 estimator with confidence threshold of 40% does lose something, as like the pretest estimator it rejects the judgmental allocation in February 2009. However, unlike the pretest estimator which goes to the boundary of the 0% confidence interval (the maximum likelihood estimator), the subjective classical estimator 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 prior.[8] 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 EuroStoxx50 time series from January 1999 until December 2015. Each set

[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.

[8]There may be choices of confidence level mappings $g(\tilde{\alpha})$ which are suboptimal, in the sense that they lead to inadmissible decision rules. An example of an inadmissible decision rule is the pretest estimator, as illustrated for instance by Magnus (1998). The $g(\tilde{\alpha})$ function associated with it was presented in section 3. A study of the properties of the $g$ functions which lead to admissible decision rules is outside the scope of the present paper.

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 3 and 4, 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 5 and 6.

Figure 5 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 generating processes which are less and less in line with the judgmental allocation. Since I know the data generating 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%.

Figure 6 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 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

[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 $\alpha|x = g(\tilde{\alpha})$.

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. This is a particularly unattractive feature of the pretest estimators, as one would like the probability of doing worse than the judgmental allocation to monotonically decrease as the judgmental decision becomes less and less accurate. 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 optimally 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 the 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 interval in a frequentist setting, it is possible to establish the equivalence between Bayesian and frequentist procedures.

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 allocation becomes greater than the given confidence level. Frequentist decisions can be interpreted as decisions under ambiguity, where the decision maker chooses the action minimizing the maximum of the expected loss over a set of priors which all give the judgmental decision in case of no data, but are left unrestricted otherwise.

# Appendix — Proofs

NOTE: Whenever there is no risk of confusion, I drop the time subscript from the random variables $X_1$ and $x_1$ to simplify the notation.

**Proof of Proposition 2.2** — Start by testing whether $\tilde{a}$ is the optimal decision, that is $H_0 : -\theta + \tilde{a} = 0$. This is equivalent to testing whether the first order conditions evaluated at $\tilde{a}$ hold.

If $c_{\alpha/2} \leq -x + \tilde{a} \leq c_{1-\alpha/2}$, the decision maker cannot reject the null hypothesis that $\tilde{a}$ is optimal and therefore selects $\tilde{a}$ as the optimal decision given the sample realization $x$.

Suppose now that $-x+\tilde{a} < c_{\alpha/2}$ and $\tilde{\tilde{a}} \equiv \delta^A(x|-x+\tilde{a} < c_{\alpha/2}) \neq x+c_{\alpha/2}$.

If $\tilde{\tilde{a}} < x + c_{\alpha/2}$, this implies that $H_0 : -\theta + \tilde{\tilde{a}} = 0$ is rejected and the expected loss can be decreased by marginal moves away from $\delta^A(x|-x+\tilde{a} < c_{\alpha/2})$. Therefore this decision rule cannot be otpimal.

If $\tilde{\tilde{a}} > x+c_{\alpha/2}$, this implies that it exists $\epsilon > 0$ such that $-x+[\tilde{\tilde{a}}-\epsilon] > c_{\alpha/2}$ and $H_0 : -\theta + [\tilde{\tilde{a}} - \epsilon] = 0$ was rejected, which implies a contradiction.

Similar arguments hold for the case $-x + \tilde{a} > c_{1-\alpha/2}$. □

**Proof of Proposition 3.1** — Note that $c_{\tilde{\alpha}} = -x + \tilde{a}$. Therefore $\delta^{A|x_1}(x) = x + c_{\tilde{\alpha}}h = x + (-x + \tilde{a})h$ and the result follows. □

**Proof of Proposition 3.2** — Impose that (8) 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 $\alpha|x_1$ gives the result. □

**Proof of Proposition 4.1** — Let's find out first the values of $a$ for which $L(\theta, a) > L(\theta, \tilde{a})$. This is equivalent to find 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:

$$\begin{aligned} P_\theta[L(\theta, \delta^{A|x_1}(X)) > L(\theta, \tilde{a})] = P_\theta[\delta^{A|x_1}(X) < \theta - |-\theta + \tilde{a}|] + \\ + P_\theta[\delta^{A|x_1}(X) > \theta + |-\theta + \tilde{a}|] \end{aligned}$$

For any $\bar{\alpha} \geq 0$ and $\underline{\alpha} \geq 0$, such that $\bar{\alpha} + \underline{\alpha} = \alpha|x_1$, I show that $P_\theta[\delta^{A|x_1}(X) < \theta - |-\theta + \tilde{a}|] \leq \bar{\alpha}$ and that $P_\theta[\delta^{A|x_1}(X) > \theta + |-\theta + \tilde{a}|] \leq \underline{\alpha}$ for all values of $\theta$ and $\tilde{a}$, with the strict inequality holding when $\theta \neq \tilde{\alpha}$.
Suppose first that $-\theta + \tilde{a} > 0$.

$$\begin{aligned} P_\theta[\delta^{A|x_1}(X) < \theta - |-\theta + \tilde{a}|] &= P_\theta[\delta^{A|x_1}(X) < 2\theta - \tilde{a}] \\ &= P_\theta[(X + c_{\underline{\alpha}})I(-X + \tilde{a} < c_{\underline{\alpha}}) + \\ &\quad + \tilde{a}I(c_{\underline{\alpha}} \leq -X + \tilde{a} \leq c_{1-\bar{\alpha}}) + \\ &\quad + (X + c_{1-\bar{\alpha}})I(-X + \tilde{a} > c_{1-\bar{\alpha}}) \\ &\quad\quad < 2\theta - \tilde{a}] \end{aligned}$$

Since the indicator functions partition $\Re$, we can bring one $\theta$ to the left, and rearranging the terms in the indicators functions gives:

$$\begin{aligned} &= P_\theta[(X + c_{\underline{\alpha}} - \theta)I(X + c_{\underline{\alpha}} - \theta > -\theta + \tilde{a}) + \\ &\quad + (-\theta + \tilde{a})I(c_{\underline{\alpha}} \leq -X + \tilde{a} \leq c_{1-\bar{\alpha}}) + \\ &\quad + (X + c_{1-\bar{\alpha}} - \theta)I(X + c_{1-\bar{\alpha}} - \theta < -\theta + \tilde{a}) \\ &\quad\quad < -(-\theta + \tilde{a})] \end{aligned}$$

Since $-\theta + \tilde{a} > 0$, the probability that the terms in the first two rows are less than $\theta - \tilde{a}$ is zero. The above probability is therefore equal to:

$$\begin{aligned} &= P_\theta[X + c_{1-\bar{\alpha}} - \theta < -(-\theta + \tilde{a})] \\ &= P_\theta[-X + \theta > c_{1-\bar{\alpha}} + (-\theta + \tilde{a})] \\ &< P_\theta[-X + \theta > c_{1-\bar{\alpha}}] \\ &= \bar{\alpha} \end{aligned}$$

where the last inequality exploits the fact that $-\theta + \tilde{a} > 0$.

Consider the other term of the probability:

$$\begin{aligned}P_\theta[\delta^{A|x_1}(X) > \theta + |-\theta + \tilde{a}|] &= P_\theta[\delta^{A|x_1}(X) > \tilde{a}] \\ &= P_\theta[(X + c_{\underline{\alpha}})I(-X + \tilde{a} < c_{\underline{\alpha}}) + \\ &\quad + \tilde{a}I(c_{\underline{\alpha}} \leq -X + \tilde{a} \leq c_{1-\bar{\alpha}}) + \\ &\quad + (X + c_{1-\bar{\alpha}})I(-X + \tilde{a} > c_{1-\bar{\alpha}}) \\ &\qquad > \tilde{a}]\end{aligned}$$

Rearranging the terms in the indicator functions:

$$\begin{aligned}&= P_\theta[(X + c_{\underline{\alpha}})I(X + c_{\underline{\alpha}} > \tilde{a}) + \\ &\quad + \tilde{a}I(c_{\underline{\alpha}} \leq -X + \tilde{a} \leq c_{1-\bar{\alpha}}) + \\ &\quad + (X + c_{1-\bar{\alpha}})I(X + c_{1-\bar{\alpha}} < \tilde{a}) \\ &\qquad > \tilde{a}]\end{aligned}$$

Since the probability that the last two rows are greater than $\tilde{a}$ is zero:

$$\begin{aligned}&= P_\theta[X + c_{\underline{\alpha}} > \tilde{a}] \\ &= P_\theta[X + c_{\underline{\alpha}} - \theta > -\theta + \tilde{a}] \\ &= P_\theta[-X + \theta < c_{\underline{\alpha}} - (-\theta + \tilde{a})] \\ &< P_\theta[-X + \theta < c_{\underline{\alpha}}] \\ &= \underline{\alpha}\end{aligned}$$

where again the last inequality exploits the fact that $-\theta + \tilde{a} > 0$.

Suppose now that $-\theta + \tilde{a} < 0$. Substituing the decision rule and rearranging the terms in the indicator functions as done in the previous case:

$$\begin{aligned}P_\theta[\delta^{A|x_1}(X) < \theta - |-\theta + \tilde{a}|] &= P_\theta[\delta^{A|x_1}(X) < \tilde{a}] \\ &= P_\theta[(X + c_{\underline{\alpha}})I(X + c_{\underline{\alpha}} > \tilde{a}) + \\ &\quad + \tilde{a}I(c_{\underline{\alpha}} \leq -X + \tilde{a} \leq c_{1-\bar{\alpha}}) + \\ &\quad + (X + c_{1-\bar{\alpha}})I(X + c_{1-\bar{\alpha}} < \tilde{a}) \\ &\qquad < \tilde{a}]\end{aligned}$$

Since the probability that the terms in the first two rows are smaller than $\tilde{a}$

is zero:

$$
\begin{aligned}
&= P_\theta[X + c_{1-\tilde{\alpha}} < \tilde{a}] \\
&= P_\theta[-X + \theta > c_{1-\tilde{\alpha}} - (-\theta + \tilde{a})] \\
&< P_\theta[-X + \theta > c_{1-\tilde{\alpha}}] \\
&= \underline{\alpha}
\end{aligned}
$$

and the inequality follows from the fact that in this case $(-\theta + \tilde{a}) < 0$.

Finally, considering the other term in the probability, after having rearranged the terms in the indicator function:

$$
\begin{aligned}
P_\theta[\delta^{A|x_1}(X) > \theta + |-\theta + \tilde{a}|] &= P_\theta[\delta^{A|x_1}(X) > 2\theta - \tilde{a}] \\
&= P_\theta[(X + c_{\underline{\alpha}})I(-X + \tilde{a} < c_{\underline{\alpha}}) + \\
&\quad + \tilde{a}I(c_{\underline{\alpha}} \leq -X + \tilde{a} \leq c_{1-\tilde{\alpha}}) + \\
&\quad + (X + c_{1-\tilde{\alpha}})I(-X + \tilde{a} < c_{1-\tilde{\alpha}}) \\
&\qquad > 2\theta - \tilde{a}]
\end{aligned}
$$

As before, since the indicator functions partition $\Re$, we can bring one $\theta$ to the left, and rearranging the terms in the indicator functions:

$$
\begin{aligned}
&= P_\theta[(X + c_{\underline{\alpha}} - \theta)I(X + c_{\underline{\alpha}} - \theta > -\theta + \tilde{a}) + \\
&\quad + (-\theta + \tilde{a})I(c_{\underline{\alpha}} \leq -X + \tilde{a} \leq c_{1-\tilde{\alpha}}) + \\
&\quad + (X + c_{1-\tilde{\alpha}} - \theta)I(X + c_{1-\tilde{\alpha}} - \theta < -\theta + \tilde{a}) \\
&\qquad > -(-\theta + \tilde{a})]
\end{aligned}
$$

Since $-\theta + \tilde{a} < 0$, the probability that the terms in the last two rows are greater than $\theta - \tilde{a}$ is zero, and the probability is equal to:

$$
\begin{aligned}
&= P_\theta[X + c_{\underline{\alpha}} - \theta > -(-\theta + \tilde{a})] \\
&= P_\theta[-X + \theta < c_{\underline{\alpha}} + (-\theta + \tilde{a})] \\
&< P_\theta[-X + \theta < c_{\underline{\alpha}}] \\
&= \underline{\alpha}
\end{aligned}
$$

where the last inequality exploits the fact that $\theta - \tilde{a} > 0$. $\square$

**Proof of Proposition 4.2** — Let's start by noting that for any decision rule, the loss function reaches its maximum when $\theta = \tilde{a}$: when the judgmental decision coincides with the optimal decision, there is no decision rule that can improve on it.

Substituting the decision rule $\delta^{A|x_1}(\bar{\alpha}; X)$ and recognizing that the choice variable is now $\bar{\alpha}$ (or alternatively $\underline{\alpha}$), the optimal decision solves the following constrained optimization problem:

$$\min_{\bar{\alpha}} E_\theta(-\delta^{A|x_1}(\bar{\alpha}; X)\theta + 0.5\delta^{A|x_1}(\bar{\alpha}; X)^2|\theta = \tilde{a})$$

Since the loss function is strictly concave, finding the $\bar{\alpha}$ that sets the first derivative equal to zero is a necessary and sufficient condition for solving the minimization problem. Imposing the condition that $\theta = \tilde{a}$ and substituting the expression for $\delta^{A|x_1}(\bar{\alpha}; X)$:

$$\begin{aligned}\min_{\bar{\alpha}} E_{\tilde{a}}(-\tilde{a}[(X + c_{\underline{\alpha}})I(-X + \tilde{a} < c_{\underline{\alpha}}) + \tilde{a}I(c_{\underline{\alpha}} < -X + \tilde{a} < c_{1-\bar{\alpha}})+\\ + (X + c_{1-\bar{\alpha}})I(-X + \tilde{a} > c_{1-\bar{\alpha}})]+\\ + 0.5[(X + c_{\underline{\alpha}})I(-X + \tilde{a} < c_{\underline{\alpha}}) + \tilde{a}I(c_{\underline{\alpha}} < -X + \tilde{a} < c_{1-\bar{\alpha}})+\\ + (X + c_{1-\bar{\alpha}})I(-X + \tilde{a} > c_{1-\bar{\alpha}})]^2)\end{aligned}$$

Let's consider each piece, after having applied a change of variable $Y = -X + \tilde{a}$. Taking the first derivative with respect to $\bar{\alpha}$ and applying Leibnitz's rule:

$$\begin{aligned}&\nabla_{\bar{\alpha}} \int_{-\infty}^{c_{\underline{\alpha}}} [(-y + \tilde{a} + c_{\underline{\alpha}})\tilde{a} - 0.5(-y + \tilde{a} + c_{\underline{\alpha}})^2] d\Phi(y) =\\ &= (\tilde{a}^2 - 0.5\tilde{a}^2)\phi(c_{\underline{\alpha}})\nabla_{\bar{\alpha}} c_{\underline{\alpha}} + \nabla_{\bar{\alpha}} c_{\underline{\alpha}} \int_{-\infty}^{c_{\underline{\alpha}}} (\tilde{a} + y - \tilde{a} - c_{\underline{\alpha}}) d\Phi(y)\end{aligned}$$

where $\phi(\cdot)$ is the pdf of the standard Normal distribution. Application of the Implicit Function Theorem gives $\nabla_{\bar{\alpha}} c_{\underline{\alpha}} = -\phi(c_{\underline{\alpha}})^{-1}$. The integral can be evaluated by applying the formula for the expectation of a truncated Normal distribution and is equal to $-c_{\underline{\alpha}}\Phi(c_{\underline{\alpha}}) - \phi(c_{\underline{\alpha}})/\Phi(c_{\underline{\alpha}})$. Putting these results together gives:

$$= -0.5\tilde{a}^2 + c_{\underline{\alpha}}\Phi(c_{\underline{\alpha}})/\phi(c_{\underline{\alpha}}) + 1/\Phi(c_{\underline{\alpha}})$$

Consider the second piece of the objective function:

$$\begin{aligned}&\nabla_{\bar{\alpha}} \int_{c_{\underline{\alpha}}}^{c_{1-\bar{\alpha}}} 0.5\tilde{a}^2 d\Phi(y) =\\ &= 0.5\tilde{a}^2\phi(c_{1-\bar{\alpha}})\nabla_{\bar{\alpha}} c_{1-\bar{\alpha}} - 0.5\tilde{a}^2\phi(c_{\underline{\alpha}})\nabla_{\bar{\alpha}} c_{\underline{\alpha}}\\ &= 0\end{aligned}$$

because application again of the Implicit Function Theorem gives $\nabla_{\bar{\alpha}} c_{\bar{\alpha}} = -\phi(c_{1-\bar{\alpha}})^{-1}$.

Consider the last piece of the objective function:

$$\begin{aligned}
&\nabla_{\bar{\alpha}} \int_{c_{1-\bar{\alpha}}}^{\infty} [(-y + \tilde{a} + c_{1-\bar{\alpha}})\tilde{a} - 0.5(-y + \tilde{a} + c_{1-\bar{\alpha}})^2] d\Phi(y) = \\
&\quad = -0.5\tilde{a}^2 \phi(c_{1-\bar{\alpha}}) \nabla_{\bar{\alpha}} c_{1-\bar{\alpha}} + \nabla_{\bar{\alpha}} c_{1-\bar{\alpha}} \int_{c_{1-\bar{\alpha}}}^{\infty} (\tilde{a} + y - \tilde{a} - c_{1-\bar{\alpha}}) d\Phi(y) = \\
&\quad = 0.5\tilde{a}^2 - \phi(c_{1-\bar{\alpha}})^{-1}[-c_{1-\bar{\alpha}}(1 - \Phi(c_{1-\bar{\alpha}})) + \phi(c_{1-\bar{\alpha}})/(1 - \Phi(c_{1-\bar{\alpha}}))] \\
&\quad = 0.5\tilde{a}^2 + c_{1-\bar{\alpha}}(1 - \Phi(c_{1-\bar{\alpha}}))/\phi(c_{1-\bar{\alpha}}) - 1/(1 - \Phi(c_{1-\bar{\alpha}}))
\end{aligned}$$

Putting the three expressions together, the first derivative is:

$$c_{\underline{\alpha}} \Phi(c_{\underline{\alpha}})/\phi(c_{\underline{\alpha}}) + 1/\Phi(c_{\underline{\alpha}}) + c_{1-\bar{\alpha}}(1 - \Phi(c_{1-\bar{\alpha}}))/\phi(c_{1-\bar{\alpha}}) - 1/(1 - \Phi(c_{1-\bar{\alpha}}))$$

which is equal to zero for $\bar{\alpha} = \underline{\alpha} = \alpha|x_1/2$. $\square$

**Proof of Proposition 4.3**

Consider all the decision rules $\delta$ of the form (12). The minimax decision rule has maximum risk $R(\tilde{a}, \alpha|x_1/2)$. By definition of $\Gamma$ there is a prior $\mu^*(\theta)$ associated with this decision rule such that:

$$\begin{aligned}
\delta^{A|x_1}(\alpha|x_1/2; x_1) &= \\
= \arg\min_{\delta} \int_X L(\tilde{a}, \delta) d\Phi(x|\theta) &= \\
= \arg\min_{\delta} \int_\Theta \int_X L(\theta, \delta) d\Phi(x|\theta) d\mu^*(\theta) &= \\
= \arg\min_{\delta} \int_\Theta \int_X L(\theta, \delta) d\mu^*(\theta|x) dm(x) &= \\
= \arg\min_{\delta} \int_X \Big[ \int_\Theta L(\theta, \delta) d\mu^*(\theta|x) \Big] dm(x) &= \\
= \arg\min_{a} \max_{\mu(\theta) \in \Gamma(A|x_1)} \int_\Theta L(\theta, a) d\mu(\theta|x)&
\end{aligned}$$

applying Fubini theorem to reverse the order of the integrals. $\square$

# 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.

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.

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. (1986), Comment (on McNees, 1986) *Journal of Business and Economic Statistics*, 4, 16-17.

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

Hansen, L. P. and T. J. Sargent (2008), *Robustness*, Princeton University Press.

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

Magnus, J. R. (1998), The Traditional Pretest Estimator, *Theory of Probability and Its Applications*, 44 (2), 293-308.

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.

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

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

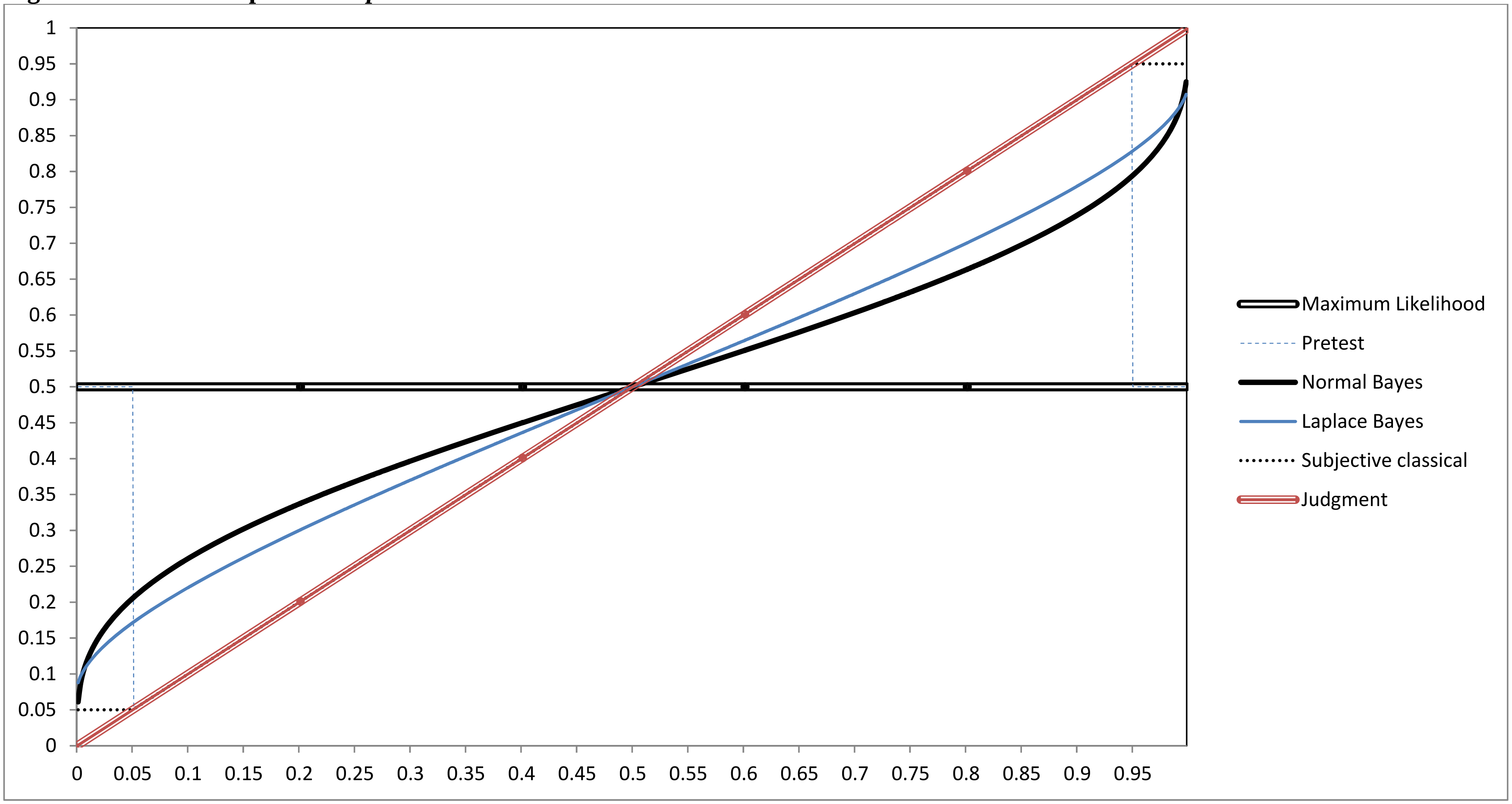
Savage, L.J. (1954), *The Foundations of Statistics*, New York, John Wiley & Sons.

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.

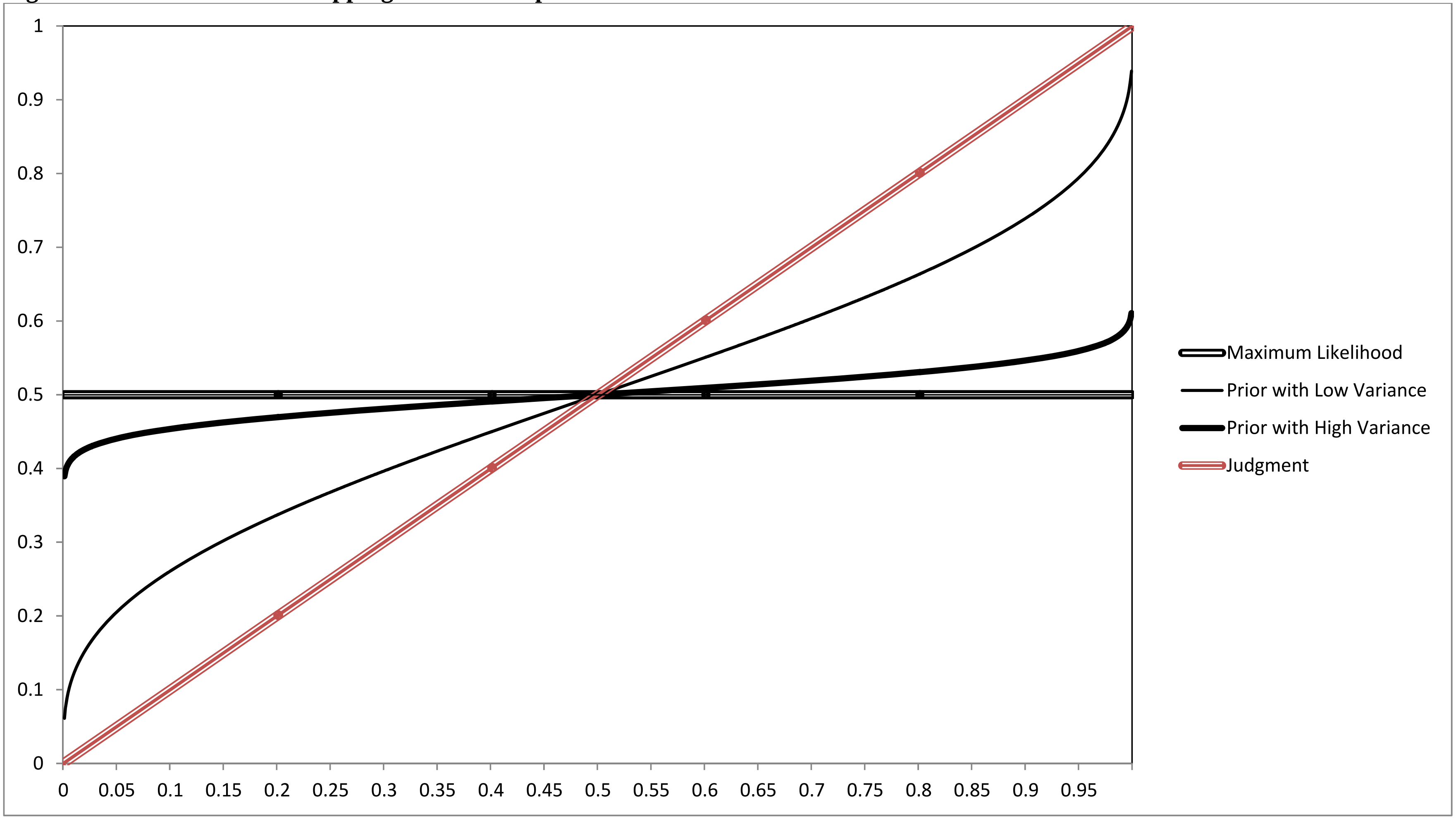
White, H. (1994), *Estimation, Inference and Specification Analysis*, Cambridge Books, Cambridge University Press.

**Figure 1 – Relationship between *p-values* and confidence levels**



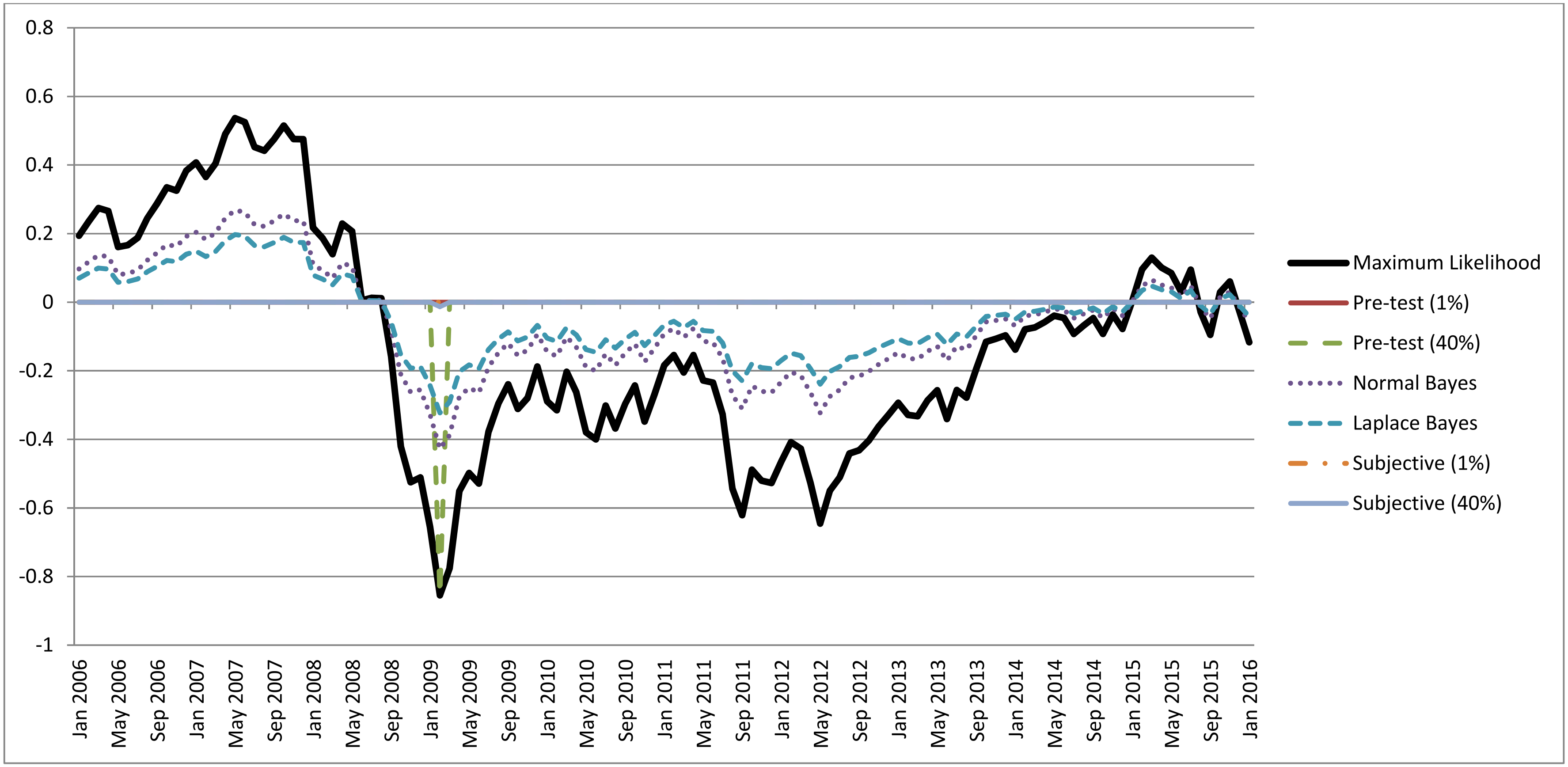

*Note*: The horizontal axis reports the *p-value* ($\tilde{\alpha}$) of the gradient $\nabla_a L(\theta, a)$ evaluated at $\theta = x_1$ and $a = \tilde{\alpha}$, the judgmental decision. The vertical axis is the chosen confidence level $\alpha | x_1 = g(\tilde{\alpha})$. The figure plots the mapping corresponding to six alternative estimators. Pre-test 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.

**Figure 2 – Confidence level mappings for Normal priors with different variances**



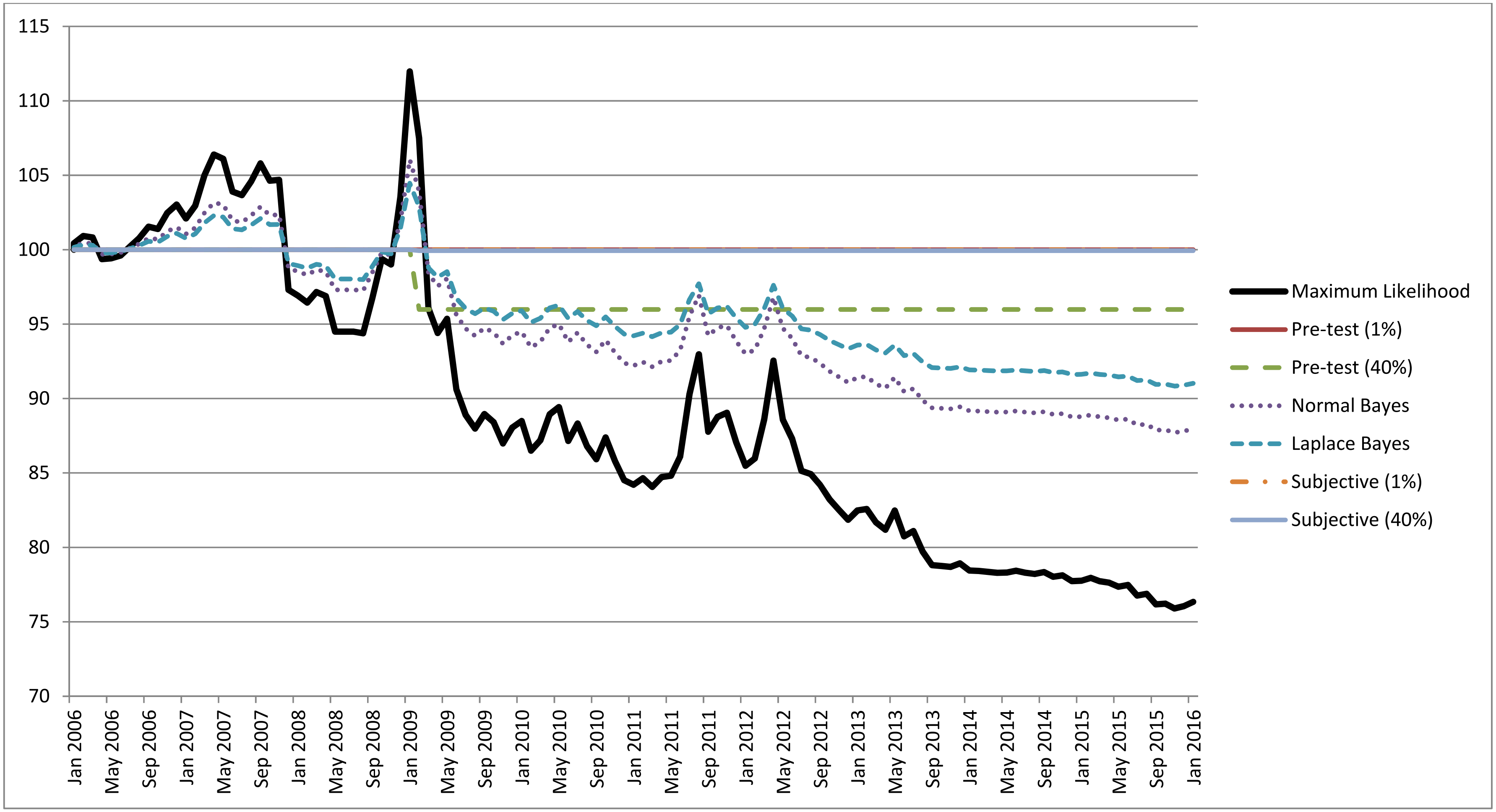

*Note*: The horizontal axis reports the *p-value* ($\tilde{\alpha}$) of the gradient $\nabla_a L(\theta, a)$ evaluated at $\theta = x_1$ and $a = \tilde{a}$, the judgmental decision. The vertical axis is the confidence level $\alpha|x_1 = g(\tilde{\alpha})$. The figure plots the confidence level mappings corresponding to Bayesian decisions based on Normal priors with mean zero and different variances.

**Figure 3 – Optimal portfolio weights**



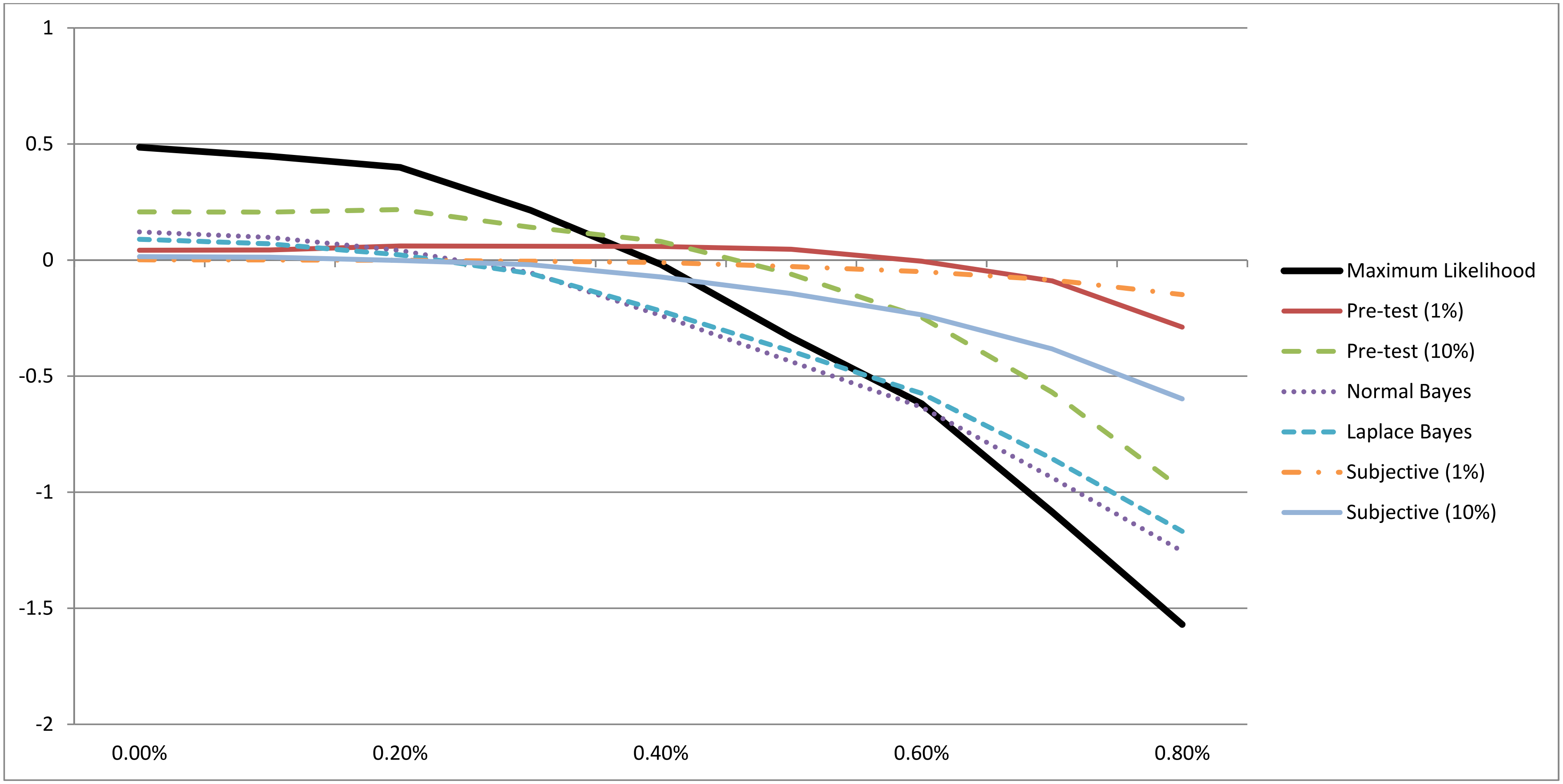

*Note*: Optimal weights according to the different decision rules of an investor choosing between cash and the EuroStoxx50 index. Weights are re-estimated each month by expanding the estimation window by one data point. The first 7 years – from January 1999 until December 2005 – are used to produce the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

**Figure 4 – Evolution of portfolio values**



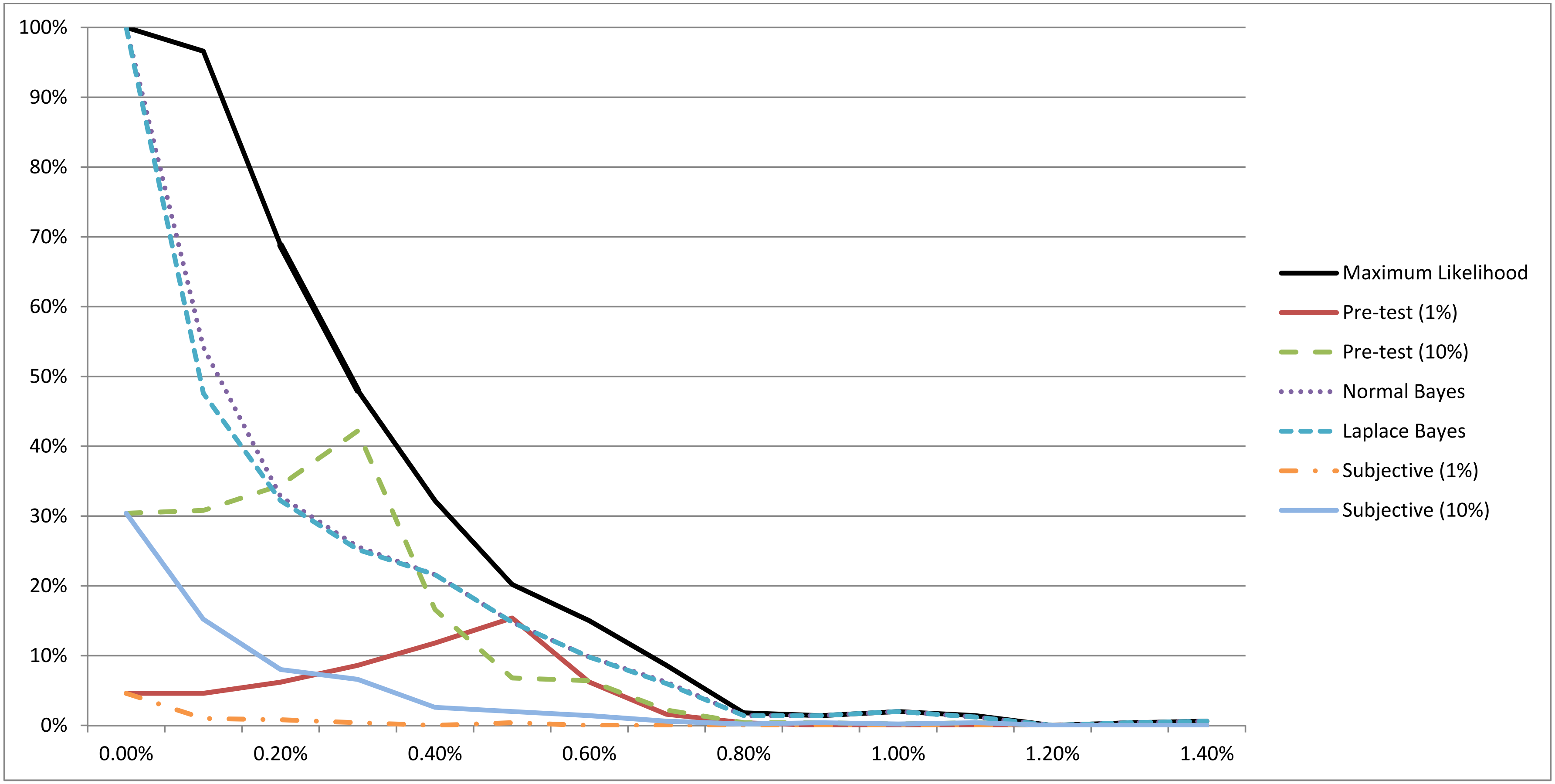

*Note*: Time evolution of the value of a portfolio invested in cash and the EuroStoxx50 index following the investment recommendations of the different decision rules.

**Figure 5 – Expected losses under alternative Data Generating 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 6 – Percentage of times statistical decisions underperform the judgmental decision for alternative Data Generating 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.