Asset Allocation with Judgment

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

June 2016

Abstract

This paper shows how to incorporate judgment in a decision problem under uncertainty, within a classical framework. The method relies on the specification of a judgmental decision with associated confidence level and application of hypothesis testing. The null hypothesis tests whether marginal deviations from the judgmental decision generate negative changes in expected utility. The resulting estimator is always at the boundary of the confidence interval: beyond that point the probability of decreasing the expected utility becomes greater than the chosen confidence level. The decision maker chooses the confidence level as a mapping from the *p-value* of the judgmental decision into the unit interval. I show how the choice of priors in Bayesian estimators is equivalent to the choice of this confidence level mapping. I illustrate the implications of this new framework with a portfolio choice between cash and the EuroStoxx50 index.

Keywords: Judgmental estimation; Statistical Risk Propensity; Out-of-sample portfolio selection.
JEL Codes: C1; C11; C12; C13; D81.

*I would like to thank for useful comments and suggestions Bruno Biais, Joel Sobel and seminar participants at the ECB and Humboldt University Berlin.

1 Introduction

The workhorse for asset allocation is the mean-variance model introduced by Markowitz (1952). Despite its success, the empirical implementation has left much to be desired. It is well-known that plug-in estimators of the portfolio weights produce very volatile asset allocations which usually perform very poorly out of sample, due to estimation errors. There is a vast literature documenting the empirical failures of the mean-variance model and suggesting possible fixes (Jobson and Korkie 1981, Michaud 1998, Brandt 2007). DeMiguel, Garlappi and Uppal (2009), however, provide convincing evidence that none of the existing solutions consistently outperforms a simple, non-statistically driven equal weight portfolio allocation.

DeMiguel et al. (2009) raise an important point: How to improve on a given judgmental allocation? Investors often approach the asset allocation problem with some default investment decision. The equal weight portfolio is one example. Other examples 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). Markowitz's model gives general prescriptions for asset allocation, but there is no guarantee that its practical implementation will out-perform a given judgmental allocation. In fact, the existing empirical evidence points to the contrary.

In this paper, I provide a theoretical framework to incorporate judgment in Markowitz portfolio allocation problems, arriving at frequentist estimators which are equivalent to Bayesian estimators, in the sense that they prescribe identical allocations for any sample realization. Judgment is summarized by a subjective allocation and a confidence level associated with it, supplied exogenously by the decision maker. It plays a role similar to the prior in Bayesian methods, even though the updating mechanism remains frequentist. Manganelli (2009) constructs the test statistic given by the first derivatives of the empirical expected utility evaluated at the judgmental allocation. If the judgmental allocation is optimal, it maximizes the true expected utility and its empirical first derivative should not be statistically different from zero. The confidence level controls the probability of committing Type I errors. Rejection of the null hypothesis implies that by moving from the judgmental allocation towards the maximum likelihood the decision maker marginally increases her expected utility in a statistically precise sense. This reasoning holds until the boundary of the confidence interval is reached, as beyond this threshold the probability of committing Type I errors becomes greater than the decision maker's confidence level. The resulting estimator is a shrinkage from the judgmental allocation to the maximum likelihood. This is the basic insight of Manganelli (2009).

The missing element needed to establish equivalence with Bayesian estimators is the choice of the confidence level. Manganelli (2009) assumes that the confidence level is set at some constant level $\bar{\alpha}$. The more general set-up proposed in this paper allows for the confidence level to depend on the evidence provided by the data, as summarized by the *p-value* of the test statistic evaluated at the judgmental allocation. For instance, a very low *p-value* may convince the decision maker to have little confidence in her initial judgment. Pre-test estimators are an extreme case in point. *P-values* higher than the confidence level induce the decision maker to stick to her judgment. *P-values* lower than the confidence level (even infinitesimally lower) induce the decision maker to neglect her judgment and adopt instead the maximum likelihood estimator.

In general, the choice of the confidence level is a mapping from the *p-value* of the test statistic evaluated at the judgmental allocation to the unit interval. The resulting confidence level α determines the width of the confidence interval and therefore the amount of shrinkage towards the maximum likelihood estimator for any given judgmental allocation and associated *p-value*. The shrinkage factor is zero if $\alpha =$ *p-value* and it is one if $\alpha = 1$. By moving continuously α over this interval, it is possible to generate any convex combination between the judgmental allocation and the maximum likelihood estimator. Therefore, any Bayesian decision shrinking from the prior to the maximum likelihood is associated with a unique mapping from the *p-value* to α in the classical framework which delivers exactly the same numerical estimate.

The proposed method satisfies also the fundamental Bayesian principle of conditioning decisions only on known variables. Judgmental decisions are based on the initial judgment, the associated confidence level (which is a function of the data) and the data itself. In particular, it does not condition on the unknown, true mean, as typically done in frequentist approaches (see for instance the discussion in section 2.4.1 in Geweke and Whiteman, 2006).

In the example considered in this paper, I derive analytically the confidence mapping associated with Bayesian estimators with Normal and Laplace priors, with zero mean and unit variance. This exercise raises two empirical challenges for Bayesian econometrics. First, Bayesian estimators with similar priors are associated with different confidence mappings. It seems an unrealistically heavy burden for the decision maker to be able to distinguish between a thin and a fat tail prior with identical means and variances, as in the case of the Normal and Laplace priors used in the example. Yet, as also confirmed by the risk analyses of Magnus (2002) and Manganelli (2009), this choice may bring very different practical implications. Second, the confidence mapping associated with the Bayesian estimators reveals that as the judgment associated with the prior becomes worse and worse, the confidence of the decision maker in her judgment increases. This result points to the risk of unintended consequences in the choice of priors in a Bayesian set-up: different priors imply different willingness to accept Type I errors in classical hypothesis testing.

The methodology is applied to a simple asset allocation problem of an investor who holds €100 and has to decide how much to invest in an Exchange Trading Fund replicating the EuroStoxx50 index. Several estimators are implemented and compared, using an out of sample exercise. The results confirm that it is difficult to find allocations with good out of sample performance. The weight associated with the maximum likelihood estimator is the most volatile. By the end of the sample, an investor who would have followed this investment strategy would have lost about one quarter of her initial wealth. The Bayesian estimators perform slightly better, as the weights are shrunk towards zero, but would still have lost between 5% and 10%. The best performing estimators 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 estimators relative to others: like investors with different risk aversion in their utility function, investors with different statistical risk propensity - that is with different willingness to tolerate Type I errors - choose different allocations. A Monte Carlo exercise shows that estimators with lower statistical risk propensity 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.[1]

[1]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).

The paper is structured as follows. The next section introduces a minimalistic asset allocation problem, which will be used as working example of the decision problem throughout the paper. Section 3 reformulates the estimation procedure proposed by Manganelli (2009). Section 4 describes the choice of the confidence level and introduces the concept of Statistical Risk Propensity. Section 5 establishes the equivalence between frequentist and Bayesian estimators with two concrete examples. Section 6 presents the empirical evidence and section 7 concludes.

2 The Asset Allocation Problem

Consider the situation of an investor holding cash and having to decide how much to invest in the stock market, for instance via an Exchange Trading Fund. Denote with $Y \sim N(\theta^0, 1)$ the return of the stock market. I assume for simplicity that stock market returns are normally distributed around an unknown mean θ^0, with known variance equal to 1. Let a denote the share of cash to be invested in the stock market. Since the nominal return on cash is zero, the portfolio return is $Z = aY$.

Assume that the investor has a mean-variance utility function (Markowitz, 1952). Her problem is to find the asset allocation a^0 which maximizes her expected utility:

$$\begin{aligned} \max_a E[U(a)] &= \max_a \{E[Z] - 0.5V[Z]\} \\ &= \max_a \{a\theta^0 - 0.5a^2\} \qquad (1) \end{aligned}$$

The first order condition of this decision problem is $\theta^0 - a = 0$. If θ^0 is known, the problem is solved by $a^0 = \theta^0$. In practice, θ^0 is unknown and needs to be estimated.

3 Econometric Solution to the Asset Allocation Problem

I assume that the econometrician has at her disposal the following elements to solve the problem:

1. A single realization from the return distribution, y_1;
2. A judgmental decision $\tilde{a}$ supplied by the decision maker;
3. A confidence level $\alpha \in [0, 1]$, describing the confidence of the decision maker in the judgmental decision $\tilde{a}$.

Building on Manganelli (2009), this section shows that the solution to the asset allocation problem requires the construction of a test statistic and of a suitable hypothesis test. The econometric solution is given by the boundary of the confidence interval of the test statistic.

3.1 Test Statistic

In the first order condition, the econometrician replaces a with the default decision $\tilde{a}$ and θ^0 with the sample mean $\hat{\theta} = Y_1$.[2] Therefore, the sample approximation of the first order condition, incorporating the judgmental decision, is the following test statistic:

$$Y_1 - \tilde{a} \tag{2}$$

Its sample realization will be different from zero with probability 1. It could be set to zero, by setting the decision a equal to y_1, as done in standard plug-in procedures. However, the interest here lies in setting to zero the first order conditions evaluated at the true parameter θ^0. Although this is unattainable in finite samples, the econometrician can test whether by moving her decision towards the maximum likelihood estimate y_1 she is more or less likely to generate positive changes in her expected utility.

3.2 Hypothesis Testing

Suppose without loss of generality that y_1 is less than $\tilde{a}$. This implies that the realization of the first derivative (2) of the maximization problem (1) evaluated at $\tilde{a}$ is negative: lower values of $\tilde{a}$, i.e. moves towards the maximum likelihood estimate y_1, generate positive changes in the empirical expected utility. The decision maker would like to rule out that the opposite is true in population, because her objective, recall, is not to increase the empirical

[2] I denote random variables with upper case letters, Y, and their realisations with lower case letters, y.

expected utility, but the population expected utility. Whenever the empirical and population expected utilities evaluated at the judgmental decision have opposite sign, deviating from her initial judgment would make her worse off. The null and alternative hypotheses which reflect this reasoning are:

$$H_0 : \theta^0 - \tilde{a} \geq 0 \qquad H_1 : \theta^0 - \tilde{a} < 0 \tag{3}$$

Under the null hypothesis, the *p-value* associated with $y_1 - \tilde{a}$ is $\tilde{\alpha} \equiv \Phi(y_1 - \tilde{a}) = P(Y_1 - \tilde{a} \leq y_1 - \tilde{a}|H_0)$, where Φ denotes the standard normal cdf. Denoting with α the confidence level chosen by the investor, if $\tilde{\alpha} < \alpha$, H_0 is rejected and the decision maker can be confident that by moving from $\tilde{a}$ towards y_1 she generates positive changes in her expected utility. This reasoning holds until the boundary of the confidence interval associated with α is reached. Defining the lower bound of the confidence interval as $\hat{a} \equiv y_1 - \Phi^{-1}(\alpha)$, any decision $a < \hat{a}$ would not be consistent with the statistical preferences of the decision maker, because for these values of a the probability of rejecting the null hypothesis when true is by construction greater than the chosen confidence level α. $\hat{a}$ is therefore 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 α. Note the economic interpretation of this type of error: the decision maker increases the sample approximation of the expected utility by moving from $\tilde{a}$ toward y_1, but in fact decreases the expected utility in population. Alternatively, she can fail to reject the null hypothesis when it is false (Type II error). This happens with probability $1 - \beta(a^0)$, where $\beta(a^0)$ is the power of the test. Economic interpretation: the decision maker could have increased the expected utility 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 $\beta(a^0)$ for values of a close to a^0. Therefore, a smaller probability of Type I errors results in a greater probability of Type II errors. It is up to the decision maker to decide how to solve this trade-off. The reasoning is summarized in table 1.

As argued by Manganelli (2009), the optimal decision in this set up is either the judgmental decision, in case the null hypothesis cannot be rejected, or the boundary of the confidence interval otherwise. It is possible to generalize this result and claim that the optimal decision is always at the boundary of the confidence interval. To arrive at this result, it is first necessary to

Table 1: Hypothesis testing

		Decision: H_0	Decision: H_1
Truth	H_0	Avoid decrease EU $1 - \alpha$	Decrease EU α
Truth	H_1	Fail to increase EU $1 - \beta(a^0)$	Increase EU $\beta(a^0)$

Note: The null hypothesis tests whether the gradient of the expected utility in population has opposite sign with respect to the sample gradient. The alternative hypothesis is that the gradient in population has the same sign as the sample gradient.
α and $\beta(a^0)$ are the size and power of the test.

discuss the choice of the confidence level associated with the hypothesis test, which is where I turn next.

4 Choosing the Confidence Level

The confidence level is typically chosen to be a constant, low number, to control the probability of committing Type I errors. There is nothing, however, that prevents the decision maker to condition her choice of α to the evidence provided by the data. A data sample which is largely inconsistent with the default decision may shatter any confidence the decision maker has in it (this is indeed what happens with the pre-test estimator). Similarly, a data sample consistent with the default decision may reinforce her confidence in it.

The confidence level may generally be considered as a mapping from the *p-value* of the test statistic (2) evaluated at y_1 into the interval $[0, 1]$. Going back to my asset allocation problem, define $\tilde{\alpha}/2 \equiv \Phi(-|y_1 - \tilde{a}|)$, where Φ is the cdf of the standard normal. $\tilde{\alpha}$ is therefore the *p-value* associated with the two-sided test statistic (2) evaluated at y_1. I define the confidence level as:

$$\alpha = g(\tilde{\alpha}) : [0,1] \to [0,1] \tag{4}$$

Note that $\alpha \geq \tilde{\alpha}$, because any decision associated with $\alpha \leq \tilde{\alpha}$ is equivalent to the decision associated with $\alpha = \tilde{\alpha}$. To see why, it is enough to notice that for any $\alpha \leq \tilde{\alpha}$ the judgmental decision $\tilde{a}$ is never rejected, because the *p-value* is greater than the chosen confidence level. In this sense, the optimal decision discussed at the end of the previous section is always at the boundary of the confidence interval, because the proposed definition of confidence level includes the special case of the boundary coinciding with the judgmental decision (this occurs when $\alpha = \tilde{\alpha}$).

Here are some common examples of how the function $g(\cdot)$ in (4) is chosen:

1. Maximum likelihood:
$$\alpha = 1,\ \forall\ \tilde{\alpha} \in [0,1]$$
2. Pre-test estimator with confidence level $\bar{\alpha}$:
$$\alpha = \begin{cases} 1 & \text{if } \tilde{\alpha} < \bar{\alpha} \\ \tilde{\alpha} & \text{if } \tilde{\alpha} \geq \bar{\alpha} \end{cases}$$
3. Subjective classical estimator with confidence level $\bar{\alpha}$ (Manganelli, 2009):
$$\alpha = \begin{cases} \bar{\alpha} & \text{if } \tilde{\alpha} < \bar{\alpha} \\ \tilde{\alpha} & \text{if } \tilde{\alpha} \geq \bar{\alpha} \end{cases}$$
4. Judgmental decision:
$$\alpha = \tilde{\alpha},\ \forall\ \tilde{\alpha} \in [0,1]$$

These examples clarify how my framework encompasses the most common estimators. The maximum likelihood estimator always disregards any default decision, by setting the confidence level equal to 1.[3] The pre-test estimator maintains the confidence $\tilde{\alpha}$ if the realization of the test statistic does not fall in the rejection region, but the confidence level is increased to 1 otherwise. The subjective classical estimator maintains the same confidence level $\bar{\alpha}$, if

[3] Note that higher values of the confidence level α actually imply lower confidence in the judgmental decision $\tilde{a}$.

the null hypothesis is rejected and otherwise it is equal to $\tilde{\alpha}$. The default decision is a special case of an estimator which never rejects the null hypothesis, by setting the confidence level always equal to $\tilde{\alpha}$.

In general, the choice of the confidence level can be any function of the *p-value* of the test statistic, as will be shown in the next section.

4.1 Statistical Risk Propensity

It may be useful to pause briefly on the economic interpretation of the choice of the confidence level. One obvious interpretation is that the confidence level reflects the confidence of the decision maker in her judgmental decision. To quote from a prominent statistician, "*in a true decision problem, the size [of the test] should be chosen according to 'subjective' factors. This is, of course, precisely what the Bayesian approach does*" (Berger 1985, p. 165). The more confident she is in her judgment the closer α will be to $\tilde{\alpha}$. In the limit, a decision maker who is absolutely certain about her judgment will set the distance between α and $\tilde{\alpha}$ equal to 0, as shown by the judgmental decision in the example above. At the other extreme, a decision maker who has no confidence at all in her judgment will maximize the distance between α and $\tilde{\alpha}$ by setting $\alpha = 1$, reverting to the maximum likelihood estimator.

There is a second appealing interpretation, which refers to the degree of statistical risk propensity, which can be loosely defined as willingness to take statistical risk. A decision maker who is very concerned about committing Type I errors, will choose α very close to $\tilde{\alpha}$, behaving like a person who is very confident in her judgment. This may be the case of a portfolio manager whose performance is evaluated by her ability to beat some benchmark portfolio, and faces severe penalties in case of underperformance relative to the benchmark. An example of such behavior is the portfolio manager of the foreign reserves of a central bank. Notice that this concept of statistical risk propensity is distinct from the standard concept of risk aversion, as summarized by the weight given to the portfolio variance in the expected utility maximization (1).

These considerations suggest the following definition of Statistical Risk Propensity (SRP):

$$SRP = 2\int_0^1 [g(\tilde{\alpha}) - \tilde{\alpha}]d\tilde{\alpha} \tag{5}$$

The minimum statistical risk propensity is reached when the decision maker sticks to her default decision, disregarding any statistical evidence. In this case, $SRP = 0$. The maximum statistical risk propensity is reached when the decision maker adopts the maximum likelihood decision, disregarding her default decision. In this case, $SRP = 1$.

5 Equivalence between Classical and Bayesian Estimators

To derive the analytical expression for the estimator, let Φ denote as usual the cdf of the standard normal distribution and suppose $y_1 > \tilde{a}$ (similar arguments go through if $y_1 < \tilde{a}$ but with different signs). The *p-value* of the test statistic evaluated at the judgmental decision is $\tilde{\alpha}/2 = \Phi[-(y_1 - \tilde{a})]$. The confidence level of the decision maker is $\alpha = g(\tilde{\alpha})$. As argued before, the judgmental estimator is given by the boundary of the confidence interval, that is $-(y_1 - \hat{a}) = \Phi^{-1}(\alpha/2)$. Define the following shrinkage factor:

$$h \equiv \frac{\Phi^{-1}(\alpha/2)}{\Phi^{-1}(\tilde{\alpha}/2)} \tag{6}$$

Note that h is always between 0 and 1, because $\alpha \geq \tilde{\alpha}$. Simple manipulations give $y_1 - \hat{a} = h(y_1 - \tilde{a})$, from which solving for $\hat{a}$:

$$\hat{a} = (1 - h)y_1 + h\tilde{a} \tag{7}$$

This estimator is a convex combination between the original judgment of the decision maker and the maximum likelihood estimator. The amount of shrinkage is determined by the factor h, which is a combination of data (as represented by y_1) and judgmental information (as represented by the judgmental decision $\tilde{a}$ and the associated confidence level α).

A Bayesian estimator maps from the prior and the data into a decision which maximizes expected utility using the posterior. In the examples analysed in this paper, the resulting decision is also a convex combination between the decision associated with the prior and the one associated with the maximum likelihood. Note that from a Bayesian perspective, the optimal allocation solves the following problem:

$$\max_a \{E[aY|y_1] - 0.5V[aY|y_1]\} \tag{8}$$

which gives the optimal Bayesian allocation

$$\hat{a}^B = \frac{E(Y|y_1)}{V(Y|y_1)} \tag{9}$$

To establish the link with Bayesian estimators it is useful to express the judgmental estimator in terms of *p-values*: $\hat{a} = \tilde{a} + [\Phi^{-1}(\alpha/2) - \Phi^{-1}(\tilde{\alpha}/2)]$. The equivalence between the judgmental and Bayesian estimators is obtained by setting $\hat{a} = \hat{a}^B$. Solving this equation with respect to α gives:

$$\alpha = 2\Phi[\Phi^{-1}(\tilde{\alpha}/2) - \tilde{a} + \frac{E(Y|y_1)}{V(Y|y_1)}] \tag{10}$$

Substituting $y_1 = \tilde{a} - \Phi^{-1}(\tilde{\alpha}/2)$ in the Bayesian estimator gives the mapping between *p-value* and confidence level which makes Bayesian and judgmental estimators equivalent.

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 mean and variance of the Bayesian predictive density are:

$$E^N(Y|y_1) = (1+c)^{-1}y_1$$
$$V^N(Y|y_1) = (1+c)^{-1}(2+c)$$

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 mean and variance of the Bayesian predictive density are:

$$E^L(Y|y_1) = E(\theta|y_1)$$
$$V^L(Y|y_1) = 1 + E(\theta^2|y_1) - E(\theta|y_1)^2$$
$$E(\theta|y_1) = y_1 - \frac{1 - \exp(2cy_1)\frac{\Phi(-y_1-c)}{\Phi(y_1-c)}}{1 + \exp(2cy_1)\frac{\Phi(-y_1-c)}{\Phi(y_1-c)}} \cdot c$$
$$E(\theta^2|y_1) = 1 - \frac{2\frac{\phi(y_1-c)}{\Phi(y_1-c)} \cdot c - [(-y_1+c)^2 + \exp(2cy_1)\frac{\Phi(-y_1-c)}{\Phi(y_1-c)}(-y_1-c)^2]}{1 + \exp(2cy_1)\frac{\Phi(-y_1-c)}{\Phi(y_1-c)}}$$

where ϕ denotes the pdf of the standard normal distribution.

In figure 1, I compare the confidence levels associated with the estimators discussed in section 4 and the two Bayesian estimators above. See also Magnus (2002) and Manganelli (2009) for a detailed risk analysis of similar estimators.

The judgmental decision is described by the diagonal line in the space $(\alpha, \tilde{\alpha})$. As already discussed in section 4, any point below this diagonal line is equivalent to its vertical projection on the diagonal.

The confidence level associated with the maximum likelihood estimator does not depend on $\tilde{\alpha}$ and is always equal to one. The confidence level of the pre-test estimator is equal to that of the judgmental decision for relatively large values of $\tilde{\alpha}$, but jumps discontinuously to one for low values of $\tilde{\alpha}$. It has the feature that small changes in $\tilde{\alpha}$ may trigger abrupt changes in confidence over the judgmental decision.

The confidence level of the subjective classical estimator proposed by Manganelli (2009) avoids the discontinuity of the pre-test estimator. It is equivalent to the Burr estimator and, being kinked, it is not admissible (Magnus, 2002).

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 level associated with the two Bayesian estimators converges to zero as $\tilde{\alpha}$ goes to zero (an exception occurs when the variance of the prior tends infinity, in which case the Bayesian and maximum likelihood estimators coincide). These two specific Bayesian estimators shrink 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 degrees of statistical risk propensity, 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. To stick to the asset allocation problem discussed in this paper, we have seen how two prior distributions with same mean and standard deviation can lead to very different decisions. 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 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 Type I errors.

In the framework proposed in this paper, instead, the decision maker provides just a judgmental decision and a confidence level with which this decision is statistically evaluated. The econometrician then tests whether the decision is supported or rejected by the data. The judgmental decision and the statistical risk that the decision maker is willing to bear should be relatively easier to supply than priors on unknown statistical parameters. For instance, a household who has to decide whether to invest in the stock market could test whether holding only the risk-free asset is an optimal decision. In the case of a portfolio manager whose performance is assessed against a benchmark, the judgmental decision could be the benchmark itself. And a central banker who has to decide on interest rates could choose not to change them (or increase, or decrease them), unless the macro-econometric model suggests otherwise.

One final comment is in order to address a fundamental criticism that is leveled against non-Bayesian methods. Geweke and Whiteman (2006), for instance, write that all non-Bayesian methods violate the *principle of relevant conditioning*, which states that forecasts should always be conditional on observed relevant events. The standard classical approach to portfolio selection solves the problem $\max_a E[U(a)]|\{y_1, \theta^0 = y_1\}$. It does in fact violate this principle, because to determine the optimal allocation it conditions on knowing the true parameter θ^0. The method proposed in this paper, instead, solves the problem $\max_a E[U(a)] \; |\{y_1, \tilde{a}, \alpha\}$. It does not violate this principle, because it does not assume that θ^0 is known to arrive at a decision, but instead uses the parameters $\{\tilde{a}, \alpha\}$ and the data y_1 to test whether the judgmental decision $\tilde{a}$ can be improved upon.

6 Empirical evidence

The previous section 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. 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 $\{\tilde{y}_t\}_{t=1}^{T}$ the original time series of log returns, let σ 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:

$$\{y_t\}_{t=1}^{T_1+s} = \{\sqrt{(T_1+s)}\tilde{y}_t/\sigma\}_{t=1}^{T_1+s} \quad \text{and} \quad \bar{y}_{T_1+s} = (T_1+s)^{-1}\sum_{t=1}^{T_1+s} y_t \qquad (11)$$

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 $\bar{y}_{T_1+s}$ 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 $\bar{y}_{T_1+s}$ for each period T_1+s. The results of this exercise are reported in figures 2 and 3. Figure 2 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 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. Pre-test and subjective estimators predict an optimal weight equal to zero, as they almost never reject the judgmental decision $\tilde{a} = 0$: the data is just too noisy to suggest a significant departure from the default decision. One needs to increase the confidence level to 40% to arrive at some rejection of the null hypothesis. That is the spike observed in February 2009 for the pre-test estimator, which for that month coincides with the maximum likelihood estimator (remember that when the pre-test estimator rejects the null hypothesis it reverts back to the maximum likelihood estimator). The weight associated with the subjective estimator with 40% confidence level exhibits just a small blip, as in case of rejection it goes to the boundary of the confidence interval.

Figure 3 report 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 estimators. Suppose the starting value of the portfolio in January is €100. By the end of the sample, after 10 years, an investor using the maximum likelihood estimator would have lost one quarter of the value of her portfolio. The situation is slightly better with the Bayesian estimators, as they imply a loss of between 5% and 10%. The pre-test estimator with confidence level 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 pre-test with 1% and the subjective estimators with confidence levels at 1% and 40% – do not lose anything because they never predict deviating from the judgmental allocation of holding all the money in cash. In fact, the subjective estimator with confidence level of 40% does lose something, as like the pre-test estimator it rejects the judgmental allocation in February 2009. However, unlike the pre-test estimator which reverts to the maximum likelihood estimator, the subjective estimator only moves to the boundary of the 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 estimator is better than the other. After all, the choice of the statistical risk propensity is a personal choice, as much as the choice of the utility function. The purpose is rather to illustrate the implications of choosing different statistical risk propensities. By choosing the maximum likelihood estimator, one has no control on the statistical risk she is going to bear. With the subjective estimator, 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 two Bayesians estimators analyzed here represent an intermediate case. The case of the EuroStoxx50 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 estimators. There is no free lunch: estimators with lower statistical risk propensity produce allocations with greater protection to underperformance relative to the judgmental allocation, but also have lower upside potential.

I illustrate this intuition with a simple simulation exercise. 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 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 2 and 3, 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 4 and 5.

Figure 4 plots the average expected utility 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 correct allocation when the 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 utility. For values of the mean close to zero, the subjective estimators dominate all the others, the one with 1% confidence level being better than the one with 10% confidence level for smaller values of the mean. As the population mean increases beyond 0.2% the Bayesian estimators start to perform better than the subjective estimators. It is only when the population mean exceeds 0.5% that the maximum likelihood estimators starts to dominate the others. Not surprisingly, decisions based on higher statistical risk propensity generate relatively higher expected utility 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 0.2%.

Figure 5 qualifies the results of figure 4. It reports the percentage of times (out of the 500 replications) that the various estimators underperform the zero judgmental allocation. When the population mean is equal to zero, subjective and pre-test 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 pre-test 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 estimator, instead, does not suffer from this drawback. Finally, the maximum likelihood and Bayesian estimators all start from an underperformance rate of 100%: when the judgmental allocation coincides with population mean, allocation based on these estimators will underperform the judgmental allocation with probability one.

7 Conclusion

This paper has developed a framework to incorporate a judgmental allocation in a statistical decision making problem of portfolio choice. A judgmental allocation can be any private information about a desired asset allocation that the decision maker brings to the decision problem. Examples could be an allocation consisting of only risk free assets, an equal weight portfolio, or some benchmark portfolio against which the performance is measured. Together with the judgmental allocation, the decision maker needs to provide a confidence level, reflecting the probability of tolerating a deviation from the judgmental allocation which results in a worse (in expected utility terms) allocation. The theory is based on testing whether the empirical first order conditions of the expected utility evaluated at the judgmental allocation are statistically different from zero, for the given confidence level. If they are, the judgmental allocation can be improved upon by moving towards the maximum likelihood estimator, until the boundary of the confidence interval is reached. Beyond this boundary, the probability of generating negative changes in the expected utility in population becomes greater than the chosen confidence level. In choosing the confidence level, the decision maker chooses a mapping from the *p-value* of the first order conditions evaluated at the judgmental decision onto the interval $[0, 1]$. I show how well-known estimators such as maximum likelihood, pre-test and Bayesian estimators based on normal and Laplace priors, map into this framework. An empirical application to the EuroStoxx50 index and a simulation study illustrate the properties of the estimators associated with different choices of the confidence level mapping. The framework can be applied to other decision making problems and is not specific to the asset allocation example.

References

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

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, 19151953.

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.

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

Jobson, J. D., and Korkie, B. (1981), Estimation for Markowitz Efficient Portfolios, *Journal of the American Statistical Association*, 75, 544554.

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.

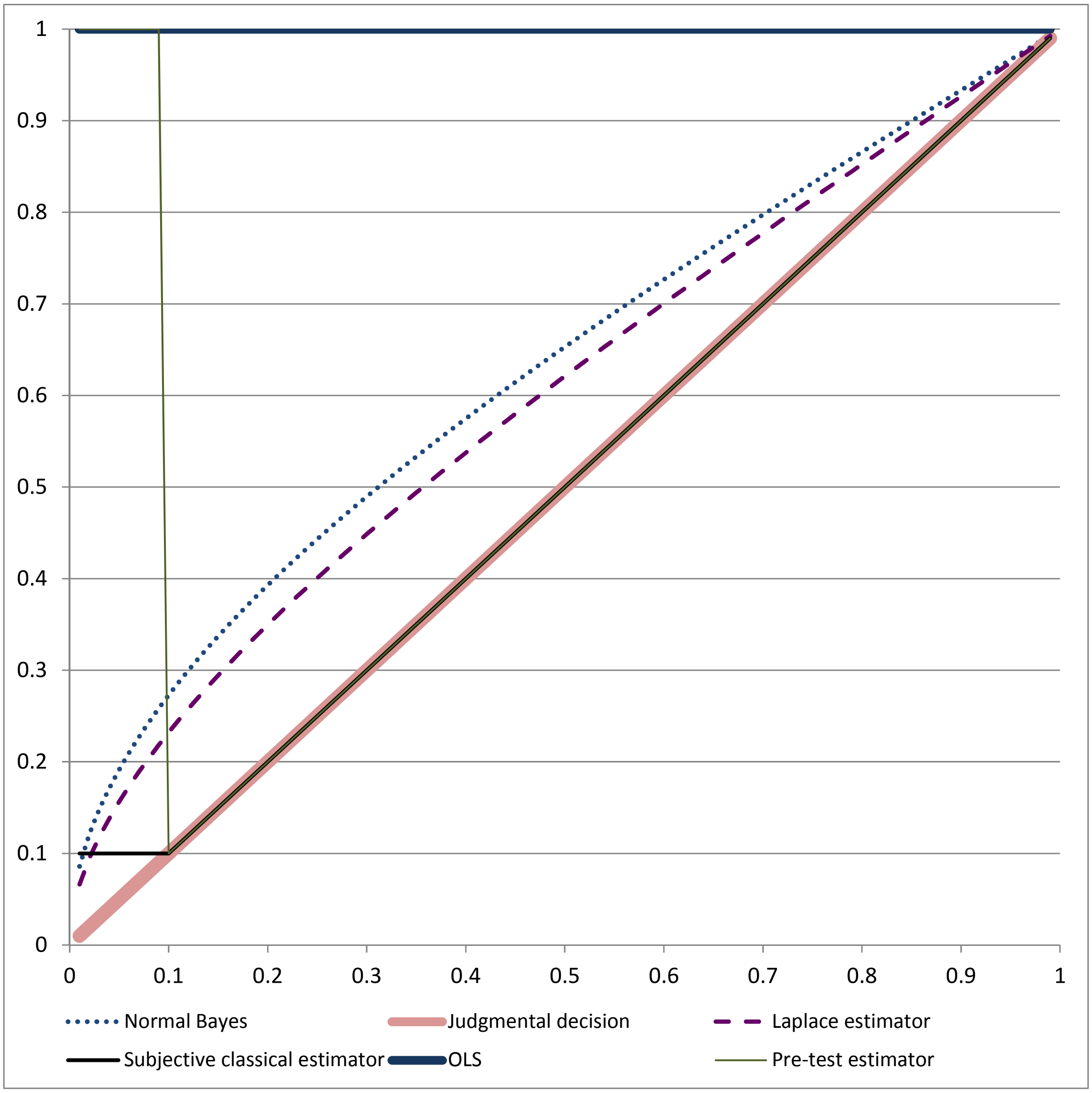
Markowitz, H. M. (1952), Portfolio Selection, *Journal of Finance*, 39, 47-61.

Michaud, R. O. (1998), *Efficient Asset Allocation. A practical Guide to Stock Portfolio Optimization and Asset Allocation*, Boston, MA: Harvard Business School Press.

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

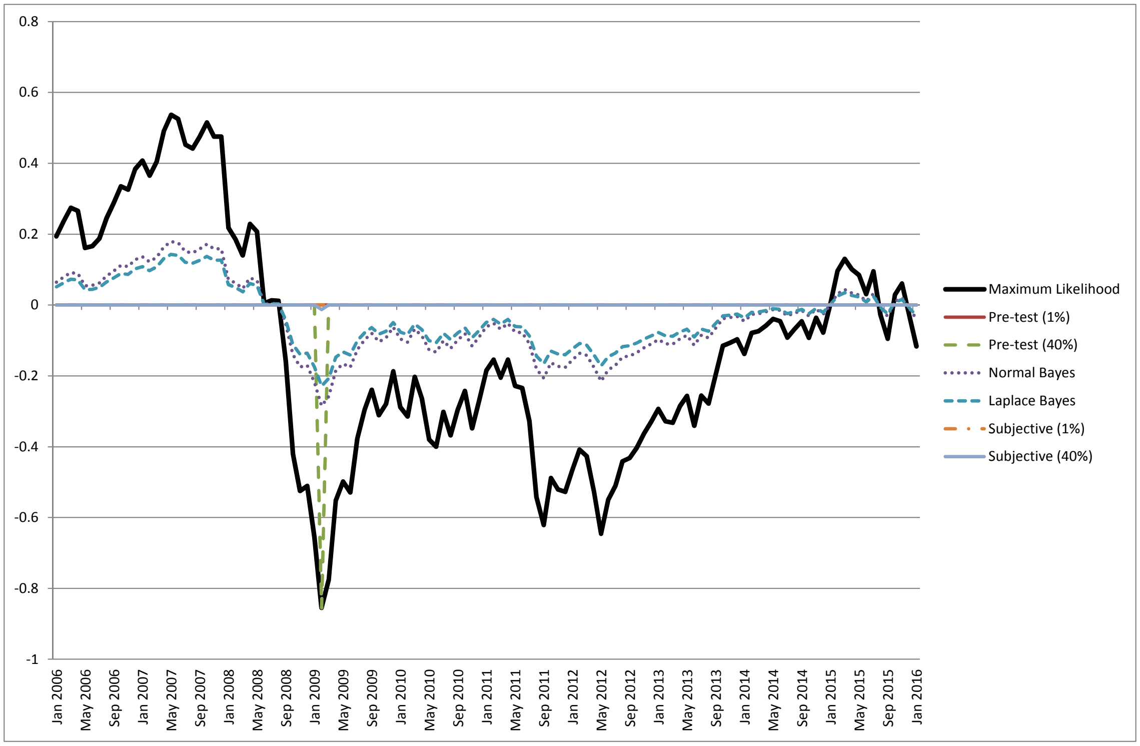
Smets, F. and R. Wouters (2007), Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach, *American Economic Review*, 97 (3): 586-606.

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



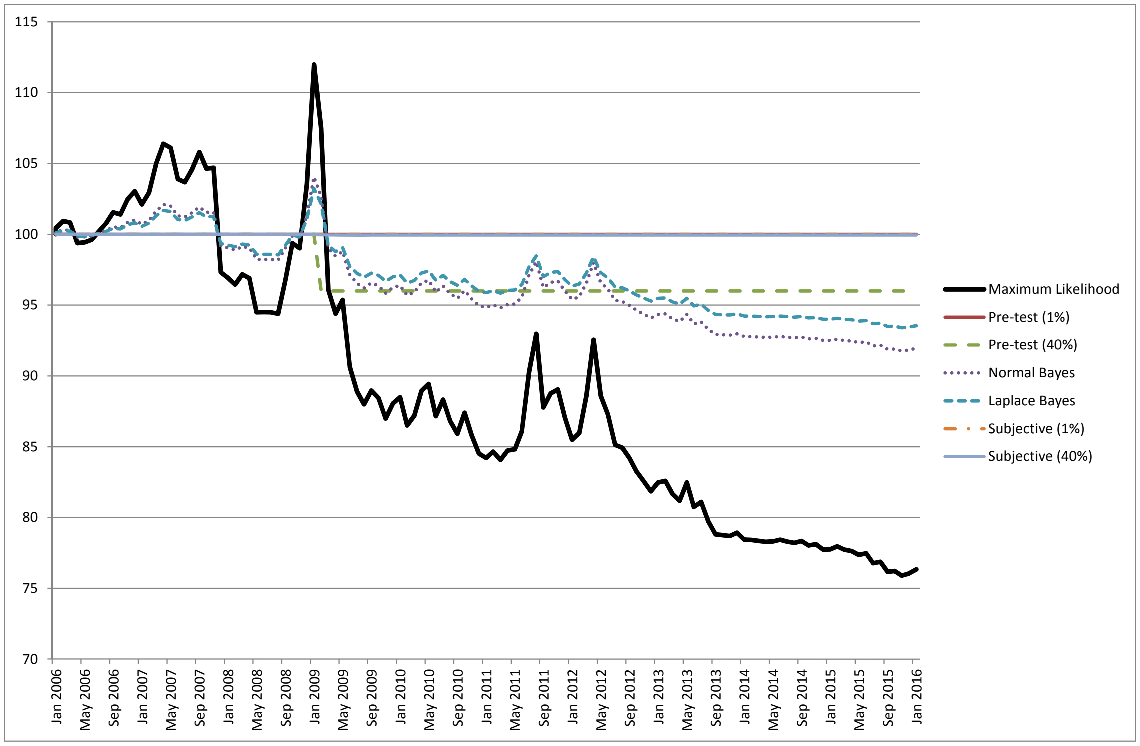
Note: The horizontal axis reports the *p-value* of the test statistic evaluated at the judgmental decision. The vertical axis is the chosen confidence level α. The figure plots the mapping corresponding to six alternative estimators. Pre-test and subjective classical estimators are based on 10% confidence levels. The normal and Laplace Bayesian estimators are based on priors with zero mean and unit variance.

Figure 2 – Optimal portfolio weights



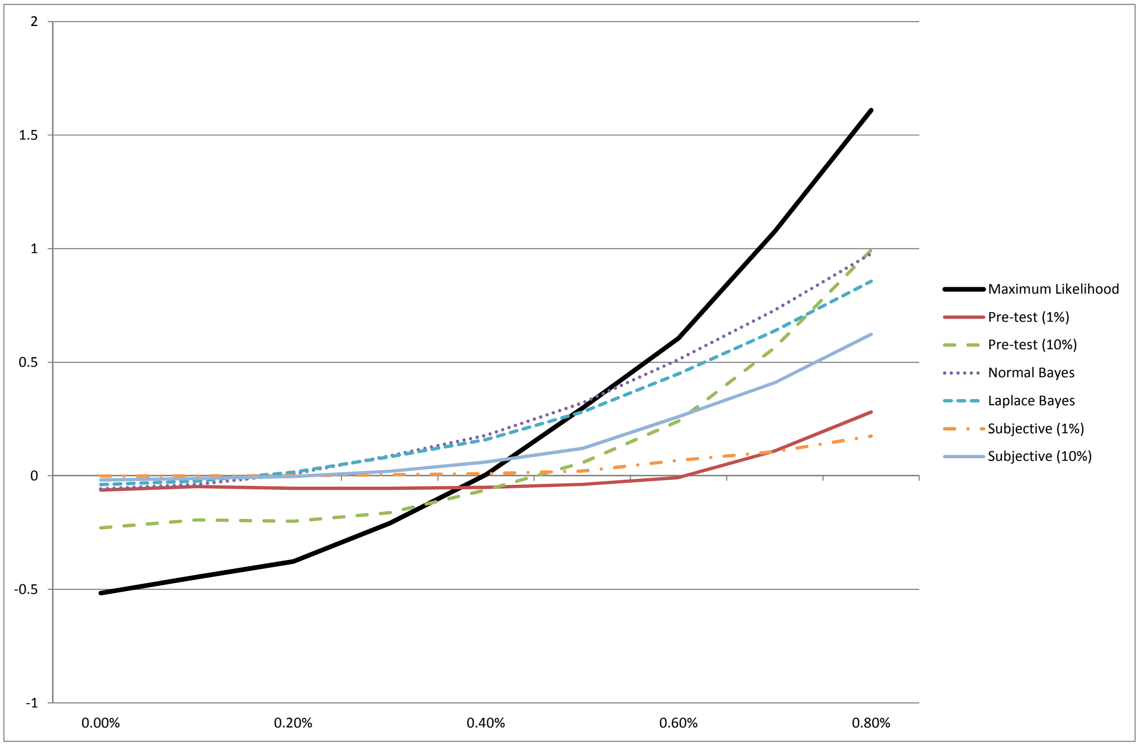
Note: Optimal weights according to the different estimators of an investor choosing between cash and the monthly EuroStoxx50 index. Weights are re-estimated each month by expanding the estimation window by one data point. The first 7 years – from January 1999 until December 2005 – are used to produce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

Figure 3 – Evolution of portfolio values



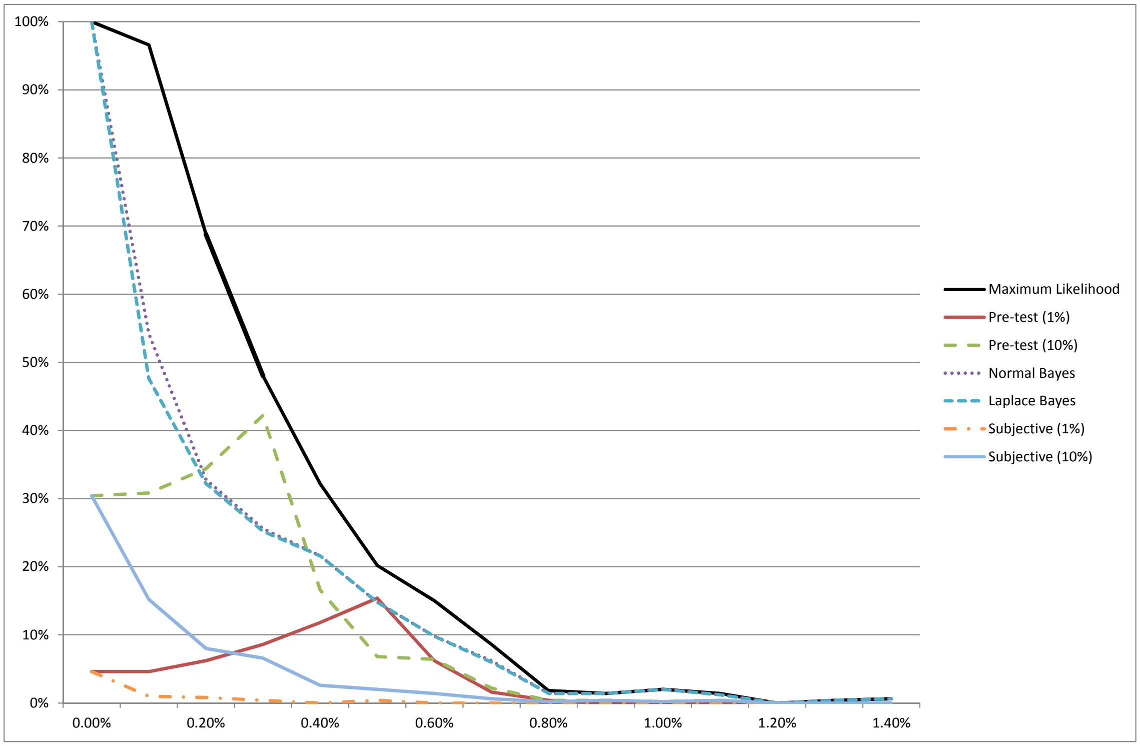
Note: Time evolution of the value of a portfolio invested in cash and the EuroStoxx50 index following the investment recommendations of the different estimators.

Figure 4 – Average Expected Utilities under alternative Data Generating Processes



Note: Average Expected Utility generated by the different estimators 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 utilities are out-of-sample averages over the 500 samples for each mean.

Figure 5 – Percentage of times the Expected Utility is negative under alternative Data Generating Processes



Note: Occurrence of negative Expected Utility generated by the different estimators under alternative specifications for the mean (reported on the horizontal axis). The simulated data are the same as in Figure 4. Negative Expected Utility implies that the estimated portfolio allocation performs worse than the judgmental allocation. Underperformance occurs more often when the judgmental allocation is close to the population mean.