Using judgment to improve benchmark asset allocations

Simone Manganelli*

March 2025

Abstract

Integrating the concept of judgment into a statistical decision framework addresses longstanding empirical challenges in asset allocation models. Judgment is defined by a benchmark allocation and a level of significance. The null hypothesis posits that the benchmark allocation is optimal if its gradient of the expected utility is zero. A test statistic evaluates the gradient using the maximum likelihood estimate. If the test does not reject the null hypothesis at the specified significance level, the benchmark allocation is retained. Otherwise, decision-makers select the allocation identified by the boundary of the confidence interval of the empirical gradient. Monte Carlo simulations and pseudo out-of-sample exercises indicate that this judgment-based approach outperforms traditional methods and the $1/N$ benchmark, providing a robust alternative for asset allocation under uncertainty.

Keywords: Asset Allocation; Statistical Decision Theory; Hypothesis Testing; Ambiguity; Asymptotics.

JEL Codes: C1; C12; C13; C18; D81.

*European Central Bank (ECB), simone.manganelli@ecb.int. The views expressed in this paper are my own and do not necessarily reflect those of the ECB.

1 Introduction

In a theoretical asset allocation problem, decision-makers choose portfolio weights that maximize a given expected utility. Markowitz (1952) provided the first such characterization, where the expected utility is represented by a trade-off between expected portfolio returns and variance. This framework paved the way for significant advancements in financial theory, including the development of the capital asset pricing model by Sharpe (1964). Despite its robust theoretical underpinnings, implementing these models empirically presents significant challenges, as highlighted by studies such as Jobson and Korkie (1980), Brandt (2009), and DeMiguel et al. (2009). Of particular concern to econometricians, DeMiguel et al. (2009) have shown that simple heuristics, like equal weight allocation, often outperform the most advanced econometric estimation techniques.

In practice, standard econometric methods often fail to provide reliable weight estimations, with plug-in estimates based on maximum likelihood tending to yield extreme and volatile allocations. These allocations frequently perform poorly out-of-sample, leading industry practitioners to adjust them using informal judgment. This paper explores how the theory of judgment developed by Manganelli (2009) and Manganelli (2025) can be applied to improve portfolio allocation by addressing these empirical challenges.

The starting point is the statistical decision theoretical framework of Wald (1950). A statistical decision problem is defined by three elements: an action, a statistical parameter and a loss function. In the context of the portfolio allocation problem explored here, the action is the asset allocation chosen by the decision-maker, the statistical parameter is the vector of expected returns and variance-covariance matrix of the portfolio assets, and the loss function is the negative of the mean-variance expected utility. Manganelli (2025) adds a fourth element, *judgment*, which is formally defined as a pair of a judgmental decision and a level of significance associated with it. The judgmental decision in the application of this paper is the benchmark allocation that the decision-maker aims to outperform.

To determine whether the benchmark allocation is optimal or if improvements are possi-

ble, we first test whether the benchmark allocation sets the gradient of the expected utility function to zero – a necessary and sufficient condition for optimality under continuous and concave losses. This test is implemented by evaluating the gradient using the maximum likelihood estimates of the statistical parameter and employing the level of significance to decide when to reject the null hypothesis. If the null hypothesis is not rejected, the benchmark allocation is retained as the final decision, as the data lack sufficient statistical evidence to warrant deviation. Conversely, if the null hypothesis is rejected, the alternative hypothesis suggests that the population gradient evaluated at the judgmental decision differs from zero, indicating that marginal adjustments in the direction of the maximum likelihood decision improve the expected utility function. This remains true until the closest boundary of the confidence interval is reached.

The frequentist approach of Manganelli (2025) introduces judgment in a statistical decision problem by imposing an external structure on the decision variables, the asset weights in the example discussed here. In contrast, the Bayesian approach incorporates judgment by imposing structure on the the model's statistical parameters. Within a simplified asset allocation example, the two approaches can be shown to be equivalent. One severe limitation of the Bayesian method is that it rejects the judgmental allocation with probability 1, offering no protection against Type I errors and resulting in dismissing an optimal judgmental allocation. This can be interpreted as the Bayesian decision-maker exhibiting uncertainty-neutral preferences, as first pointed out by Ellsberg (1961). The robust Bayesian approach, which maximizes expected utility under a worst-case prior scenario, does not automatically reject the Bayesian-equivalent of the judgmental decision, aligning more closely with the approach advocated in this paper. A limitation of the Bayesian approach is that it is unclear how the priors on statistical parameters translate to the decision variables. In asset allocation problems, it is often more realistic and intuitive to express judgment directly in terms of decision variables rather than statistical parameters.

A key ingredient in the implementation of the test is the asymptotic distribution of the maximum likelihood estimators of the mean and variance-covariance matrices of the

portfolio assets, needed to construct the distribution of the test statistic. This distribution is easily available using standard results. For this paper, we rely on the results of Lütkepohl (2005). In order to assess the accuracy of the asymptotic approximation, we perform several Monte Carlo experiments.

First, we check the size and power of the test. We simulate multivariate normal data, using the empirical mean and variance-covariance matrix of a dataset from DeMiguel et al. (2009). For each simulation, we test the null hypothesis that the gradient of the expected utility evaluated at the judgmental decision is equal to zero. Given our knowledge of the data generating process (DGP), we can evaluate the size of the test using as judgmental decision the theoretical optimal allocation. Our findings indicate that, the rejection rate aligns precisely with the significance level (set at 10%), even for small estimation windows of as few as 60 observations. This holds for different number of assets (1, 5, 20 and 50) and for various degrees of risk aversion in the expected utility.

To assess the power of the test, we employ the equal weight allocation strategy suggested by DeMiguel et al. (2009). Because the quality of the equal weight allocation is unknown *a priori*, determining the power of the test is inherently challenging. Typically, the power of a test is reduced when the tested hypothesis is close to the true hypothesis. Our analysis reveals that the test's power is quite sensitive to the length of the estimation window and, notably, to the degree of risk aversion. It is not surprising that the power increases with the sample size. The power's increase with the risk aversion coefficient is also reasonable, given that second moments are generally easier to estimate than first moments: as the risk aversion coefficient rises, decision-makers place greater emphasis on the portfolio's variance, thereby enhancing the test's accuracy.

Second, we conduct a similar Monte Carlo exercise to compare the expected utilities of the decision with judgment against those of the benchmark allocation. Again, we use as benchmark allocations the theoretically optimal and the $1/N$ portfolios. Our findings corroborate the conclusion of DeMiguel et al. (2009) that the standard plug-in estimator performs poorly out-of-sample. In contrast, the decision with judgment, which constrains

Type I errors, performs only slightly worse than the optimal allocation (which cannot be improved by construction) and significantly outperforms the $1/N$ benchmark.

Finally, we employ the same data sets of DeMiguel et al. (2009) to evaluate how the decision with judgment fares in a pseudo out-of-sample exercise with real data. Consistently with the simulation exercise, the decisions with judgment outperform the $1/N$ benchmark portfolio, except for one dataset.

The paper is structured as follows. The next section frames the asset allocation problem in a statistical decision theoretical context and explains how to arrive at the decision with judgment. Section 3 explain the connection of the decision with judgment with the Bayesian and robust Bayesian decisions, in the context of a simplified example. Section 4 contains techinical details about the implementation of the decision with judgment. Section 5 presents the Monte Carlo exercises. Section 6 applied the decision with judgment to the data set of DeMiguel et al. (2009). Section 7 concludes.

2 Asset Allocation with Judgment

2.1 The problem

Consider a portfolio consisting of $N+1$ assets, where the $(N+1)^{th}$ asset is the risk-free asset with return R. Let $\mathbf{a} \equiv [a_1, \ldots, a_N]'$ represent the N-vector of portfolio weights for the first n assets. The return of asset i at time t is denoted by x_{it}. The portfolio return, accounting for the constraint that the weights sum to one, is expressed as:

$$\begin{aligned} y_t(\mathbf{a}) &= \sum_{i=1}^{N} a_i x_{it} + \left(1 - \sum_{i=1}^{N} a_i\right) R \\ &\equiv \mathbf{a}'\mathbf{x}_t + (1 - \mathbf{a}'\iota)R, \end{aligned} \tag{1}$$

where $\mathbf{x}_t$ is the N-vector containing x_{it} for $i = 1, \ldots, N$, and ι is an N-vector of ones. Assume also that a series T of observations is observed.

Define the vector of expected returns and the variance-covariance matrix as follows:

$$\underset{(N\times 1)}{\mu} \equiv E[\mathbf{x}_t] \tag{2}$$

$$\underset{(N\times N)}{\Omega} \equiv V[\mathbf{x}_t], \tag{3}$$

Define also

$$\underset{(N(N+3)/2\times 1)}{\theta} \equiv [\mu', \text{vech}(\Omega)']', \tag{4}$$

where vech(Ω) is the operator that stacks the elements on and below the main diagonal of the variance-covariance matrix. Consider an investor with a mean-variance expected utility function:

$$U(\theta, \mathbf{a}) = \mathbf{a}'\mu + (1 - \mathbf{a}'\iota)R - 0.5\gamma\mathbf{a}'\Omega\mathbf{a}, \tag{5}$$

with $\gamma > 0$.

The asset allocation problem can be reformulated as:

$$\max_{\mathbf{a}} U(\theta, \mathbf{a}). \tag{6}$$

2.2 The solution with judgment

Manganelli (2025) defines judgment as a pair formed by a judgmental decision $\tilde{\mathbf{a}} \in \mathbb{R}^N$ and a level of significance $\alpha \in [0, 1]$:

$$A \equiv \{\tilde{\mathbf{a}}, \alpha\}. \tag{7}$$

In the context of the current asset allocation problem, let $\tilde{\mathbf{a}}$ represent a benchmark allocation that an investor aims to outperform. This allocation will also be referred to as judgmental decision or judgmental allocation. This subsection explores whether and how the given benchmark allocation can be statistically improved.

The expected utility (5) depends on the vector of statistical parameters θ that needs to be estimated. If $\hat{\theta}(\mathbf{X})$ denotes the maximum likelihood estimator, assuming that the

conditions for consistency and asymptotic normality are satisfied, its distribution is:

$$\sqrt{T}(\hat{\theta}(\mathbf{X}) - \theta) \overset{a}{\sim} N(0, \Sigma), \tag{8}$$

where $\mathbf{X}$ is a $(T \times N)$ matrix of random returns.

Define the maximum likelihood decision $\hat{\mathbf{a}}(\mathbf{x})$ associated with it:

$$\begin{aligned} \hat{\mathbf{a}}(\mathbf{x}) &= \arg\max_{\mathbf{a}} U(\hat{\theta}, \mathbf{a}) \\ &= \hat{\Omega}(\mathbf{x})^{-1}(\hat{\mu}(\mathbf{x}) - \iota R)/\gamma. \end{aligned} \tag{9}$$

where $\mathbf{x} \equiv [\mathbf{x}_1; \ldots; \mathbf{x}_T]$ is the $(T \times N)$ matrix of observed sample return.

Consider the following shrinking action from the judgmental decision $\tilde{\mathbf{a}}$ to the maximum likelihood decision:

$$\mathbf{a}(\lambda; \mathbf{x}) = \lambda \hat{\mathbf{a}}(\mathbf{x}) + (1 - \lambda)\tilde{\mathbf{a}}, \text{ for } \lambda \in [0, 1] \tag{10}$$

and rewrite the expected utility as a function of λ, which becomes the relevant decision variable. The asset allocation problem becomes:

$$\max_{\lambda} V(\theta, \lambda; \mathbf{x}) \equiv \max_{\lambda} \left[\mathbf{a}(\lambda; \mathbf{x})'\mu + (1 - \mathbf{a}(\lambda; \mathbf{x})'\iota)R - 0.5\gamma \mathbf{a}(\lambda; \mathbf{x})'\Omega \mathbf{a}(\lambda; \mathbf{x})\right]. \tag{11}$$

Two important remarks are in order. First, observe the transformation from an N-variate (as a function of $\mathbf{a}$) to a univariate (as a function of λ) optimization problem. While this transformation may result in some loss of generality, it offers significant advantages in terms of operationalization. It is theoretically possible to consider an N-variate shrinkage parameter $\boldsymbol{\lambda} \in [0, 1]^N$ (see Manganelli (2025) for a discussion). Second, the expected utility depends on the population parameter θ, but it conditions also on the sample realization $\mathbf{x}$ through $\mathbf{a}(\lambda; \mathbf{x})$. This conditioning is crucial for setting up the Monte Carlo simulation and for establishing a connection with Bayesian decisions, which will be explored in the next section.

Since $\tilde{\mathbf{a}} = \mathbf{a}(0; \mathbf{x})$, the null hypothesis that the judgmental decision $\tilde{\mathbf{a}}$ is optimal can be formulated as

$$H_0 : \nabla_\lambda V(\theta, 0; \mathbf{x}) = 0. \tag{12}$$

This can be tested by constructing the following statistic:

$$\begin{aligned}\sqrt{T}\nabla_\lambda V(\hat{\theta}(\mathbf{Y}), \lambda; \mathbf{x}) &= \sqrt{T}\nabla'_{\lambda,\theta} V(\bar{\theta}(\mathbf{Y}), \lambda; \mathbf{x})(\hat{\theta}(\mathbf{Y}) - \theta) \\ &\overset{a}{\sim} N(0, \sigma^2),\end{aligned}$$

where $\mathbf{Y}$ is an auxiliary $(T \times N)$ random matrix with the same distribution as $\mathbf{X}$, the mean value vector $\bar{\theta}(\mathbf{Y})$ lies between $\hat{\theta}(\mathbf{Y})$ and θ, the asymptotic variance is

$$\sigma^2 \equiv \nabla'_{\lambda,\theta} U(\theta, \lambda) \Sigma \nabla_{\lambda,\theta} U(\theta, \lambda) \tag{13}$$

and Σ is the variance-covariance matrix of the maximum likelihood estimator from equation (8). The auxiliary random matrix $\mathbf{Y}$ is needed to avoid confusing its potential realizations $\mathbf{y}$ with the actual realizations $\mathbf{x}$ on which the null hypothesis (12) conditions.

Since $\mathbf{Y}$ has the same distribution as $\mathbf{X}$, we can use the realization $\mathbf{x}$ to construct the following test statistic:

$$\sqrt{T}\hat{\sigma}(\mathbf{x})^{-1}\nabla_\lambda V(\hat{\theta}(\mathbf{x}), \lambda; \mathbf{x}) \tag{14}$$

Manganelli (2025) shows that the decision with judgment is then:

$$\hat{\lambda} = \max\{0, \lambda^*\}, \tag{15}$$

where λ^* is implicitly defined by

$$\sqrt{T}\hat{\sigma}(\mathbf{x})^{-1}\nabla_\lambda V(\hat{\theta}(\mathbf{x}), \lambda^*; \mathbf{x}) = c_{1-\alpha/2} \tag{16}$$

and $c_{1-\alpha/2}$ is the critical value of the standard normal distribution. The critical value is constructed using the given level of significance in equation (7), which together with the judgmental decision $\tilde{\mathbf{a}}$ forms the (exogenous) judgment of the decision-maker.

The decision (15) can be understood through the following economic intuition. If the null hypothesis (12) is not rejected, the decision-maker opts for the benchmark allocation $\tilde{\mathbf{a}}$, as there is insufficient statistical evidence to justify a deviation. Rejection of the null hypothesis, on the other hand, implies acceptance of the alternative. The alternative hypothesis states that the population gradient evaluated at the benchmark allocation is positive, indicating that incremental adjustments toward the maximum likelihood decision (9) increase the expected utility. These adjustments continue until the decision $\mathbf{a}(\lambda^*; \mathbf{x})$ is reached, marking the boundary of the confidence interval. At this point, the null hypothesis asserting the optimality of $\mathbf{a}(\lambda^*; \mathbf{x})$ is no longer rejected.

3 Connection with the Bayesian approach

Decisions with judgment condition on the sample realization, similar to Bayesian decisions (see, for instance, Section 1.6.3 of Berger (1985)). These methods offer alternative strategies for integrating judgment into the statistical decision problem, as described by equation (6). The frequentist approach advocated by Manganelli (2025) incorporates judgment by imposing an external structure on the decision variable $\mathbf{a}$. In contrast, the Bayesian approach introduces judgment by imposing structure on the the statistical parameters θ.

This section discusses the explicit link between decisions with judgment and Bayesian decisions, in the context of a simplified asset allocation problem studied in this paper.

3.1 Bayesian decision

Suppose there is only one risky asset, so that the weight $a \in \mathbb{R}$ denotes the fraction invested in that asset and $(1 - a)$ the fraction invested in cash yielding a 0 return. Assume there is only one sample realization x, that the expected returns are $\mu = E[X]$ and their variance

is $\Omega = 1$. Assume finally that the risk aversion is $\gamma = 1$. The expected utility is now $U(\mu, a) = a\mu - 0.5a^2$.

The Bayesian approach requires the decision-maker to formulate a prior on the statistical parameter μ. Let's assume, again for the sake of simplicity that the prior is $\mu \sim N(0, 1)$. To such prior corresponds the judgmental decision

$$\begin{aligned} a^{\mu} &= \arg\max_{a} \int (a\mu - 0.5a^2) d\mu \\ &= 0, \end{aligned} \tag{17}$$

that is, zero investment in the risky asset.

Assuming that the likelihood of the data is $X \sim N(0, 1)$, the posterior distribution of μ upon observing the sample realization x is $\mu|x \sim N(x/2, 1/2)$. The Bayesian decision based on the posterior is:

$$\begin{aligned} a^{\mu|x} &= \arg\max_{a} \int (a\mu - 0.5a^2) d\mu|x \\ &= x/2. \end{aligned} \tag{18}$$

After observing the sample realization, the fraction invested in the risky asset is modified in the direction of the maximum likelihood estimate x. For the prior chosen in this example, this results in an investment in the risky asset equal to $x/2$.

This decision coincides with the decision with judgment when $\tilde{a} = a^{\mu} = 0$ and $a^{\mu|x} = a(\lambda^*; x)$, which in this simple example is given by $\lambda^* = 1/2$. By equation (16), this in turn implies that the level of significance α is implicitly defined by

$$\begin{aligned} \hat{\sigma}(x)^{-1} \nabla_{\lambda} V(\hat{\theta}(x), 1/2; x) &= |x|^{-1}(x^2 - x^2/2) \\ &= |x|/2 = c_{1-\alpha/2}. \end{aligned}$$

To recap, in the simplified setting of the example considered in this section, a Bayesian

decision-maker with a standard normal prior on the statistical parameter μ behaves like a decision-maker with judgment given by:

$$\tilde{a} = 0 \text{ and } \alpha|x = 2\left(1 - \Phi(|x|/2)\right). \tag{19}$$

This stylized example raises several concerns about the Bayesian decision-making approach, both from practical and theoretical perspectives. Starting with practical challenges, the prior distribution on the model's statistical parameter implicitly imposes a specific judgment on the decision-maker, as defined in (7). In the example above, a normal prior on μ is equivalent to assuming the judgmental decision and the level of significance shown in (19). While calibrating priors to reasonable judgment might be feasible in simple cases, it becomes increasingly complex as the number of statistical parameters grows. For instance, the empirical application discussed later in the paper involves a portfolio with 50 assets, requiring the estimation of 1325 parameters, even under the simplifying assumption of constant means and variance-covariance matrix. Accounting for time-varying mean and second moments further complicates prior calibration. Furthermore, expressing judgment directly through decision variables seems to be more intuitive from a practitioner's standpoint.

The second theoretical concern is that Bayesian decisions inherently reject the judgmental decision $\tilde{a}$ implied by the prior with probability 1. This is like assuming an *ex ante* level of significance of 1, which is revised to a value less than 1 after observing the sample realization, as determined by equation (19). For a detailed discussion of this example, see Manganelli (2023), who distinguishes between *ex ante* and *ex post* significance levels. The lack of control over Type I errors is particularly problematic in asset allocation: when the benchmark portfolio is near-optimal, Bayesian decisions may unjustifiable reject it and shrink towards the maximum likelihood decision, potentially resulting in a suboptimal allocation. Conversely, decisions with judgment can use significance levels to guard against incorrectly rejecting an optimal benchmark allocation, as will be demonstrated in the Monte Carlo section. This concern is linked to the fact that Bayesian decisions are

ambiguity-neutral, as first noted by Ellsberg (1961). There is now a large literature on ambiguity aversion, including asset allocation applications such as Epstein and Schneider (2010), which highlights the theoretical limitations of the Bayesian approach. Recent reviews can be found in Gilboa and Marinacci (2013) or Ilut and Schneider (2023). The next subsection illustrates how decisions with judgment can be interpreted as ambiguity-averse decisions.

3.2 Robust Bayesian decision

Decision-makers with a level of significance set to 1 can be regarded as having uncertainty-neutral preferences. An early critique of the Bayesian approach highlighted that Bayesian decision-makers do not differentiate between scenarios with a precisely known prior and those with a range of priors for μ. To address this critique, Gilboa and Schmeidler (1989) proposed relaxing the independence axiom of Savage (1954) and replacing it with the uncertainty aversion axiom (see Gilboa (2009) for an in-depth discussion). When decision-makers are characterized by a set of priors, they maximize the expected utility under the worst-case scenario among those priors. Models incorporating ambiguity-aversion have since become the dominant paradigm in decision theory. Let us apply this framework to the simplified example discussed in this section.

Let us assume that the prior distribution on μ is now uncertain and given by a whole set of distributions:

$$\mu \sim N(\pi, 1) \qquad -1 < \pi < 1.$$

The decision for the ambiguity-averse decision-maker is expressed as:

$$\max_{a} \min_{\pi} \int (a\mu - 0.5a^2) d\mu$$

The judgmental decision based only on the priors is $a^{\mu(\pi)} = 0$. The robust Bayesian

decision, conditioned on the observed realization x, is:

$$a^{\mu(\pi)|x} = \begin{cases} -(|x|-1)/2 & \text{if } x < -1 \\ 0 & \text{if } -1 \leq x \leq 1 \\ (|x|-1)/2 & \text{if } x > 1 \end{cases}$$

Here, the judgmental decision $a^{\mu(\pi)} = 0$ is retained when the sample realization x is between -1 and 1. This corresponds to not rejecting the judgmental decision with an *ex ante* level of significance given by:

$$\alpha = 2\Phi(-1).$$

The correspondence with the decision with judgment is obtained by setting again $a^{\mu(\pi)|x} = a(\lambda^*; x)$, which gives:

$$\lambda^* = (|x|-1)/2|x|$$

Substituting λ^* in equation (16) results in:

$$\hat{\sigma}(x)^{-1}\nabla_\lambda V(\hat{\theta}(x), \lambda^*; x) = (|x|+1)/2,$$

which corresponds to the following *ex post* level of significance:

$$\alpha|x = 2\left(1 - \Phi\left(\frac{|x|+1}{2}\right)\right).$$

Unlike Bayesian decisions, robust Bayesian decisions do not automatically reject the judgmental decision, thereby offering protection against Type I errors. However, similar to the Bayesian approach, judgment is implicitly embedded in the formulation of the set of prior distributions. This complicates implementation, as eliciting prior distributions on the model's statistical parameters is more challenging than expressing judgment on the decision variables controlled by the decision-maker. The benefit of linking decisions with judgment to Bayesian and robust Bayesian decisions is that they inherit the theoretical

optimality results of these approaches.

4 Implementation of the decision with judgment

Implementation of formula (15) requires the computation of several elements. Let us proceed in order.

4.1 Maximum likelihood estimates and their asymptotic distribution

To derive the properties of the maximum likelihood estimator $\hat{\theta}(\mathbf{Y})$, rewrite the model as a simple regression:

$$\mathbf{Y}_t = \mu + \mathbf{u}_t \qquad \mathbf{u}_t \sim N(0, \Omega). \tag{20}$$

Given the matrix of sample realizations $\mathbf{x}$, the maximum likelihood estimates are (see, for instance, formulae 3.4.9 and 3.4.11 of Lütkepohl (2005)):

$$\hat{\mu}(\mathbf{x}) = T^{-1} \sum_{t=1}^{T} \mathbf{x}_t \tag{21}$$

$$\hat{\Omega}(\mathbf{x}) = T^{-1} \sum_{t=1}^{T} (\mathbf{x}_t - \hat{\mu}(\mathbf{x}))(\mathbf{x}_t - \hat{\mu}(\mathbf{x}))'. \tag{22}$$

The asymptotic distribution is given by Proposition 3.4 of Lütkepohl (2005):

$$\sqrt{T}(\hat{\theta}(\mathbf{Y}) - \theta) \overset{a}{\sim} N\left(\mathbf{0}_{(N(N+3)/2\times 1)}, \begin{bmatrix} \Omega & \mathbf{0}_{(N\times N(N+1)/2)} \\ \mathbf{0}_{(N(N+1)/2\times N)} & \Sigma_\Omega \end{bmatrix}\right), \tag{23}$$

where:

$$\Sigma_\Omega = 2\bar{D}_N(\Omega \otimes \Omega)\bar{D}_N' \tag{24}$$

$$\bar{D}_N = (D_N' D_N)^{-1} D_N' \tag{25}$$

and D_N is the $(N^2 \times N(N+1)/2)$ duplication matrix, such that $\text{vec}(\Omega) = D_N \text{vech}(\Omega)$.

4.2 Derivatives of the utility function

The other elements necessary to compute the test statistic (14) are the derivatives of the utility function.

$$\nabla_\lambda V(\theta, \lambda; \mathbf{x}) = \bar{\mathbf{a}}'(\mu - \iota R - \gamma \Omega \mathbf{a}(\lambda)), \tag{26}$$

where

$$\bar{\mathbf{a}} \equiv \hat{\mathbf{a}} - \tilde{\mathbf{a}}. \tag{27}$$

The cross-derivative of the mean value expansion is:

$$\nabla_{\lambda,\theta} V(\theta, \lambda; \mathbf{x}) = \begin{bmatrix} \bar{\mathbf{a}} \\ -\gamma D_N'(\bar{\mathbf{a}} \otimes \mathbf{a}(\lambda)) \end{bmatrix} \tag{28}$$

5 Monte Carlo evaluation

One key ingredient of the decision with judgment is the asymptotic distribution of the maximum likelihood estimator. It is therefore essential to evaluate the accuracy of the approximation. One first important aspect is the accuracy of the size of the test based on the statistic (14). The second important aspect is the power of the test, in case the judgmental decision is not optimal. The power inevitably depends on the accuracy of the initial judgmental decision. Finally, we want to understand under which circumstances the allocation with judgment outperforms the judgmental decision, as measured by the expected utility. The following two subsections review each of these issues.

5.1 Size and power of the test

The data generating process for the Monte Carlo evaluation is as follows. We utilized simulated data from DeMiguel et al. (2009), which in turn followed the methodology of MacKinlay and Pastor (2000). The data were generated using a single factor-model, where

the factor excess returns have an annual average of 8% and standard deviation of 16%. The intercept was set to zero, and the factor loadings were evenly distributed between 0.5 and 1.5. The variance-covariance matrix of the factor regression residuals was assumed to be diagonal, with elements drawn from a uniform distribution between 0.10 and 0.20. The data were generated once for 50 assets, and for exercises involving fewer assets, we simply select the first subset of assets from this dataset.

We constructed portfolios with $N = 1, 5, 20, 50$ assets, by selecting the first N assets from the above-mentioned dataset, using the mean and variance-covariance matrices of these portfolios as DGP. We generated a random draw from this DGP to construct $\hat{a}$, the maximum likelihood allocation. Next, we generated 10,000 draws from the same DGP and computed the test statistic (14) for each draw, noting whether it exceeded the critical value. This exercise was repeated for different estimation windows (60, 120, 180, ..., 600) and various judgmental decisions.

The optimal decision is given by:

$$\mathbf{a}^0 = \Omega^{-1}(\mu - \iota R)/\gamma, \tag{29}$$

where μ and Ω are the population values from the DGP. The judgmental decisions where chosen as a linear combination between the optimal decision and the $1/N$ allocation (denoted by $\mathbf{a}_N$): $\tilde{\mathbf{a}} = P\mathbf{a}^0 + (1-P)\mathbf{a}_N$, for $P = 0, 0.2, 0.4, \ldots, 1$. When $P = 1$ the judgmental allocation coincides with $\mathbf{a}^0$ and when $P = 0$ with $\mathbf{a}_N$. Intermediate values of P correspond to varying degrees of suboptimal judgmental allocations. Finally, we repeated the exercise for different values of the risk aversion coefficient, $\gamma = 1, 5, 10$. It is unclear how to assess *ex ante* the degree of suboptimality of the $1/N$ allocation. By varying the risk aversion coefficient, we can adjust how closely the $1/N$ allocation aligns with or deviates from the optimal decision $\mathbf{a}^0$.

The results are presented in Tables 1, 2, 3 and Figures 1, 2, 3 for a level of significance $\alpha = 0.10$. Each figures plots the results from the corresponding table, divided in four

quadrants representing the portfolios with $N = 1$, 5, 20 and 50 assets. Each quadrant features six lines, each corresponding to one of the six judgmental allocations considered in the exercise.

Several key findings are worth highlighting. First, the size of the test statistic (reported as dashed black line) is unbiased. When the judgmental decision is optimal ($\tilde{\mathbf{a}} = \mathbf{a}^0$), the null hypothesis (12) should be rejected approximately α of the times. The tables and the figures show that the asymptotic approximation of the test size is quite accurate, even for the smallest estimation window (60 observations) and the largest portfolios (50 assets). The probability of incorrectly rejecting the optimal judgmental decision remains around 10%, regardless of the estimation window length or degree of risk aversion.

Second, the power of the test is heavily influenced by the quality of the judgmental decision, the number of assets in the portfolio and the sample size for parameter estimation. Recall that the equally weighted judgmental portfolio represents one extreme of the judgmental allocation (the other extreme being the optimal allocation), shown as a black line. For instance, in the case of $\gamma = 1$ and 5 or 10 assets, the equally weighted allocation appears very close to the optimal one in the space determined by the test statistic (14), as all lines coincide. The situation is quite different when $\gamma = 5$ and even more so when $\gamma = 10$. In this latter case, the power against the equally weight allocation steadily increases with the estimation window size, reaching 1 already for an estimation sample of 240 observations. Naturally, the power is lower for allocations closer to the optimal decision: the closer the judgmental allocation to the optimal one, the lower the power, as is evident in Figure 3.

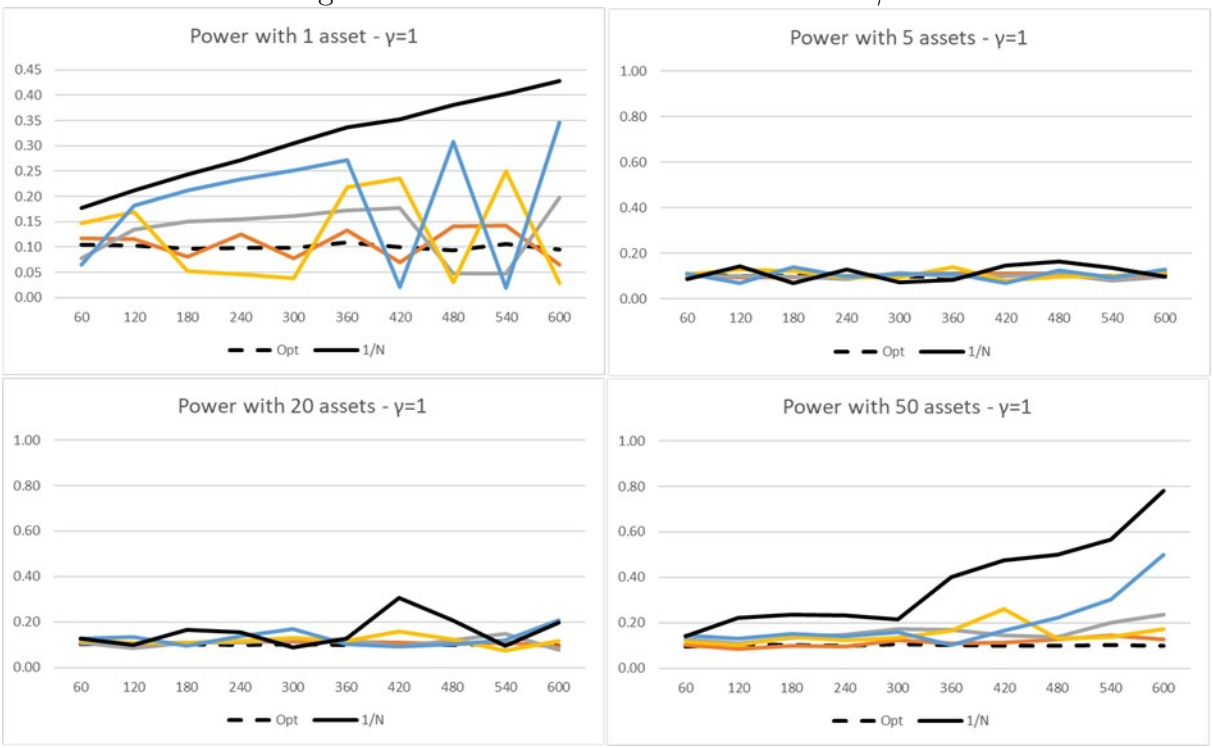
Table 1: Power of the test statistic when $\gamma = 1$

γ=1
α=0.10

# assets	ã	Window size 60	120	180	240	300	360	420	480	540	600
1	Opt	0.10	0.10	0.10	0.10	0.10	0.11	0.10	0.09	0.11	0.10
		0.12	0.12	0.08	0.13	0.08	0.13	0.07	0.14	0.14	0.06
		0.08	0.13	0.15	0.16	0.16	0.17	0.18	0.05	0.05	0.20
		0.15	0.17	0.05	0.05	0.04	0.22	0.24	0.03	0.25	0.03
		0.06	0.18	0.21	0.23	0.25	0.27	0.02	0.31	0.02	0.35
	1/N	0.18	0.21	0.24	0.27	0.31	0.34	0.35	0.38	0.40	0.43
5	Opt	0.11	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.11
		0.10	0.09	0.09	0.10	0.11	0.11	0.11	0.11	0.10	0.11
		0.10	0.10	0.10	0.09	0.11	0.10	0.10	0.11	0.08	0.10
		0.11	0.13	0.12	0.09	0.09	0.14	0.08	0.10	0.10	0.12
		0.11	0.07	0.14	0.10	0.11	0.11	0.07	0.12	0.09	0.13
	1/N	0.09	0.14	0.07	0.13	0.07	0.08	0.15	0.16	0.13	0.10
20	Opt	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.11	0.10
		0.10	0.10	0.11	0.11	0.11	0.11	0.11	0.10	0.11	0.09
		0.11	0.09	0.11	0.11	0.13	0.13	0.09	0.12	0.15	0.08
		0.12	0.11	0.11	0.12	0.13	0.12	0.16	0.12	0.08	0.12
		0.13	0.13	0.10	0.14	0.17	0.10	0.09	0.10	0.12	0.21
	1/N	0.13	0.10	0.17	0.15	0.09	0.13	0.31	0.21	0.10	0.20
50	Opt	0.10	0.11	0.10	0.10	0.11	0.10	0.10	0.10	0.10	0.10
		0.10	0.09	0.10	0.10	0.12	0.11	0.11	0.13	0.14	0.13
		0.13	0.11	0.14	0.15	0.17	0.17	0.15	0.14	0.20	0.24
		0.12	0.10	0.14	0.12	0.14	0.17	0.26	0.13	0.14	0.17
		0.15	0.13	0.15	0.14	0.16	0.10	0.17	0.22	0.30	0.50
	1/N	0.14	0.22	0.23	0.23	0.21	0.40	0.48	0.50	0.57	0.78

Note: Power for different judgmental decisions is assessed between two extremes. On one end is the optimal decision $\mathbf{a}^0$, as defined in (29). On the opposite end is the $1/N$ equal weight allocation, denoted by $\mathbf{a}_N$. The intermediate rows represent linear combinations of these two decisions, given by $\tilde{\mathbf{a}} = P\mathbf{a}^0 + (1-P)\mathbf{a}_N$, for $P = 0, 0.2, 0.4, \ldots, 1$. The different columns indicate the various window sizes used to estimate the test statistic (14), while the separate blocks of rows present the results for different number of assets included in the portfolio.

Figure 1: Power of the test statistic when $\gamma = 1$



Note: Graphical representation of the results of Table 1. Each quadrant reports the power for portfolios with varying number of assets. The lines within each quadrant represent the power for different judgmental allocations. The optimal allocation $\mathbf{a}^0$ is reported as a dashed black line, while the equal weight allocation $\mathbf{a}_N$ is represented as a continuous black lines. The lines in between correspond to various combinations of these two allocations.

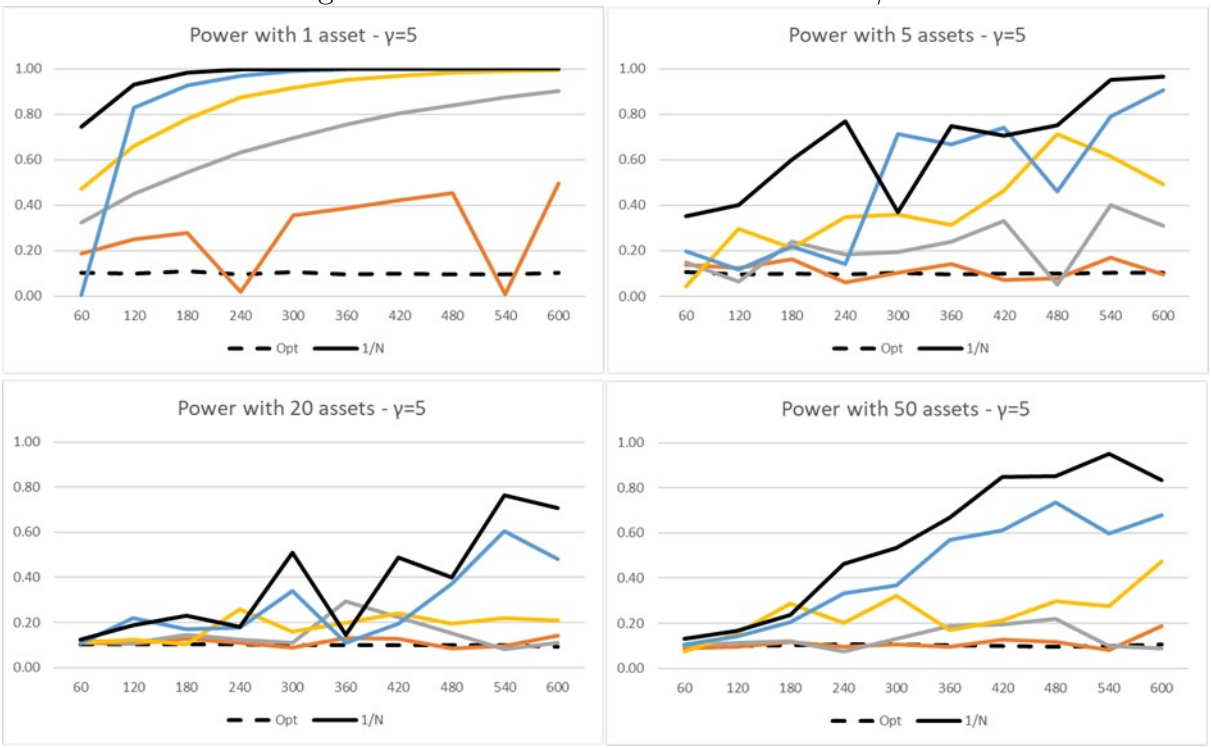
Table 2: Power of the test statistic when $\gamma = 5$

γ=5
α=0.10

# assets	ã	Window size 60	120	180	240	300	360	420	480	540	600
1	Opt	0.10	0.10	0.11	0.10	0.11	0.10	0.10	0.10	0.10	0.10
		0.19	0.25	0.28	0.02	0.36	0.39	0.42	0.45	0.01	0.49
		0.32	0.45	0.54	0.63	0.70	0.75	0.81	0.84	0.88	0.90
		0.47	0.66	0.78	0.87	0.92	0.95	0.97	0.98	0.99	0.99
		0.00	0.83	0.93	0.97	0.99	1.00	1.00	1.00	1.00	1.00
	1/N	0.74	0.93	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	Opt	0.11	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
		0.14	0.13	0.16	0.06	0.10	0.14	0.07	0.08	0.17	0.10
		0.15	0.07	0.24	0.18	0.19	0.24	0.33	0.05	0.40	0.31
		0.04	0.30	0.21	0.35	0.36	0.32	0.47	0.71	0.61	0.49
		0.20	0.12	0.22	0.14	0.71	0.67	0.74	0.46	0.79	0.91
	1/N	0.35	0.40	0.60	0.77	0.37	0.75	0.71	0.75	0.95	0.97
20	Opt	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.09
		0.11	0.11	0.13	0.11	0.09	0.13	0.13	0.09	0.09	0.14
		0.12	0.11	0.15	0.12	0.11	0.29	0.22	0.15	0.08	0.11
		0.11	0.12	0.10	0.26	0.16	0.20	0.24	0.20	0.22	0.21
		0.11	0.22	0.17	0.18	0.34	0.11	0.20	0.37	0.60	0.48
	1/N	0.13	0.19	0.23	0.18	0.51	0.14	0.49	0.40	0.76	0.71
50	Opt	0.10	0.10	0.10	0.10	0.11	0.10	0.10	0.10	0.10	0.11
		0.09	0.10	0.12	0.09	0.11	0.10	0.13	0.12	0.08	0.19
		0.10	0.11	0.12	0.07	0.13	0.19	0.19	0.22	0.10	0.09
		0.07	0.16	0.29	0.20	0.32	0.17	0.21	0.30	0.28	0.48
		0.11	0.14	0.20	0.33	0.37	0.57	0.61	0.74	0.60	0.68
	1/N	0.13	0.17	0.24	0.46	0.53	0.67	0.85	0.85	0.95	0.84

Note: See note of Table 1.

Figure 2: Power of the test statistic when $\gamma = 5$



Note: See note of Figure 1.

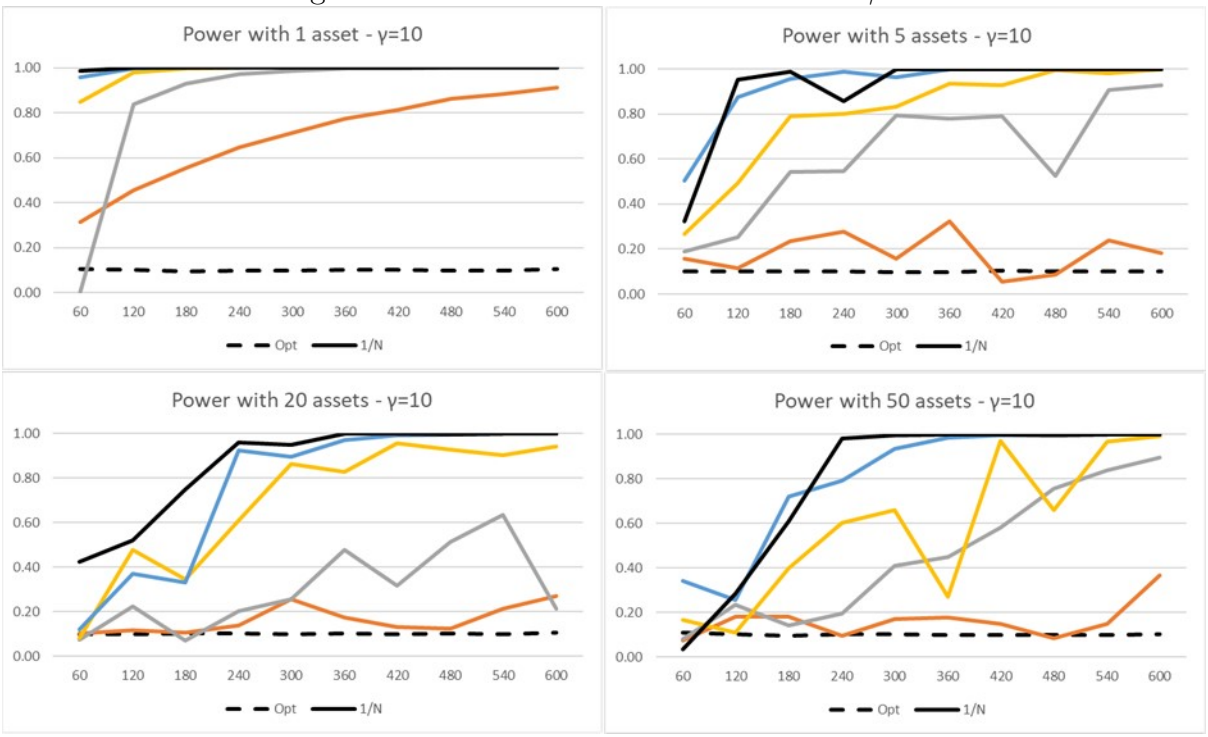
Table 3: Power of the test statistic when $\gamma = 10$

γ=10
α=0.10

# assets	ã	Window size 60	120	180	240	300	360	420	480	540	600
1	Opt	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.11
		0.32	0.46	0.56	0.65	0.71	0.78	0.81	0.86	0.88	0.91
		0.00	0.84	0.93	0.97	0.99	1.00	1.00	1.00	1.00	1.00
		0.85	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1/N	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	Opt	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
		0.16	0.11	0.24	0.28	0.16	0.32	0.05	0.09	0.24	0.18
		0.19	0.25	0.54	0.55	0.79	0.78	0.79	0.53	0.91	0.93
		0.27	0.49	0.79	0.80	0.83	0.94	0.93	1.00	0.98	1.00
		0.50	0.87	0.96	0.99	0.96	1.00	1.00	1.00	1.00	1.00
	1/N	0.32	0.95	0.99	0.86	1.00	1.00	1.00	1.00	1.00	1.00
20	Opt	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.11
		0.10	0.12	0.11	0.14	0.25	0.17	0.13	0.12	0.21	0.27
		0.07	0.22	0.07	0.20	0.26	0.48	0.32	0.51	0.63	0.21
		0.08	0.48	0.34	0.61	0.86	0.83	0.96	0.93	0.90	0.94
		0.12	0.37	0.33	0.92	0.89	0.97	0.99	1.00	1.00	1.00
	1/N	0.42	0.52	0.75	0.96	0.95	1.00	1.00	1.00	1.00	1.00
50	Opt	0.11	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
		0.08	0.18	0.18	0.10	0.17	0.18	0.15	0.09	0.15	0.37
		0.08	0.23	0.14	0.19	0.41	0.45	0.58	0.76	0.84	0.90
		0.17	0.11	0.40	0.60	0.66	0.27	0.97	0.66	0.97	0.99
		0.34	0.26	0.72	0.79	0.93	0.99	1.00	1.00	1.00	1.00
	1/N	0.04	0.28	0.61	0.98	1.00	1.00	1.00	1.00	1.00	1.00

Note: See note of Table 1.

Figure 3: Power of the test statistic when $\gamma = 10$



Note: See note of Figure 1.

5.2 Expected utility performance

The data generating process for the evaluation of the expected utility performance mirrors that described in the previous subsection. To streamline the exercise, we focused on three estimation windows, 180, 300 and 600 observations, corresponding to 15, 25 and 50 years of monthly data, respectively. We assessed two judgmental portfolios, the optimal and the equally weighted. Additionally, we considered three levels of significance, $\alpha = 1$, 0.1 and 0.01. The case of $\alpha = 1$ corresponds to the standard plug-in estimator, as in DeMiguel et al. (2009), where there is no protection against estimation error. Since the quality of the equal weight portfolio cannot be assessed *ex ante*, we examined three risk aversion coefficients, $\gamma = 1$, 5 and 10. The previous subsection's exercise indicated that the power against the equally weighted portfolio is quite low for low levels of risk aversion, suggesting that the equally weighted portfolio is not distant from the optimal one for less risk-averse investors. However, the power increases substantially for higher levels of risk aversion. A plausible explanation for this result is the well-known fact that first moments are estimated much less precisely than second moments. Since less risk-averse agents place more emphasis on expected returns relative to the variance, one would expect a substantially lower performance in terms of expected utility for these agents, compared to those with a higher degree of risk aversion.

The exercise is structured as follows. Using the same dataset, we consider portfolios comprising 1, 5, 20 and 50 assets. For each portfolio, we computed the expected returns and the variance-covariance matrix, which serve as DGP. We then generate a random draw that matches the size of the estimation window to compute the maximum likelihood allocation. Subsequently, we generate 10,000 random draws from the same GDP, each with the same length as the estimation window, and implement the decision rule (15) for each draw. Next, we compute the expected utility (using the population moments from the DGP) for both the decision with judgment and the judgmental decision. Finally, we average the expected utilities over all the 10,000 draws and report the difference between the expected utility of the decision with judgment and that of the judgmental decision.

Table 4: Expected utility

			α=1			α=0.1			α=0.01		
	# assets	ã	180	300	600	180	300	600	180	300	600
γ=1	1	Opt	-3.32	-1.81	-0.17	-0.09	-0.07	-0.005	-0.01	-0.003	0.000
		1/N	-1.05	-0.20	-1.49	-0.01	-0.03	-0.01	-0.001	-0.002	-0.001
	5	Opt	-23.82	-4.73	-7.17	-0.10	-0.02	-0.03	-0.005	-0.001	-0.001
		1/N	-25.70	-20.62	-3.87	-0.10	-0.05	-0.01	-0.01	-0.004	**0.001**
	20	Opt	-78.02	-45.05	-13.52	-0.13	-0.06	-0.03	-0.01	-0.003	-0.002
		1/N	-68.26	-50.46	-19.69	-0.10	-0.02	**0.05**	-0.004	**0.002**	**0.01**
	50	Opt	-527.46	-141.39	-43.37	-0.12	-0.07	-0.03	-0.01	-0.004	-0.001
		1/N	-466.36	-98.26	-54.51	-0.01	-0.06	**1.93**	**0.01**	-0.002	**0.73**
γ=5	1	Opt	-0.66	-3.09	-0.01	-0.02	-0.01	-0.001	-0.001	-0.001	0.000
		1/N	**5.47**	**5.35**	**5.45**	**3.98**	**4.55**	**5.03**	**2.17**	**3.23**	**4.34**
	5	Opt	-3.60	-0.05	-0.70	-0.02	-0.003	-0.01	-0.001	-0.001	0.000
		1/N	-2.17	-0.68	**2.12**	**1.09**	**0.43**	**2.16**	**0.31**	**0.80**	**2.29**
	20	Opt	-16.55	-8.53	-2.33	-0.03	-0.01	-0.01	-0.001	-0.001	-0.0003
		1/N	-19.00	-8.70	-0.06	0.06	**0.03**	**0.75**	**0.04**	**0.002**	**0.34**
	50	Opt	-61.63	-28.18	-10.47	-0.02	-0.01	-0.01	-0.002	-0.001	-0.0003
		1/N	-46.35	-25.04	-3.50	-0.02	**0.04**	**1.16**	**0.00**	**0.06**	**0.08**
γ=10	1	Opt	-0.33	-0.18	-0.02	-0.01	-0.01	-0.001	0.00	0.00	0.00
		1/N	**13.81**	**13.90**	**13.77**	**13.10**	**13.46**	**13.64**	**11.83**	**12.74**	**13.34**
	5	Opt	-2.38	-0.47	-0.72	-0.01	-0.01	-0.003	-0.001	-0.0004	-0.0002
		1/N	**9.56**	**10.06**	**11.74**	**9.15**	**10.19**	**11.61**	**7.77**	**9.42**	**11.29**
	20	Opt	-7.80	-4.50	-1.35	-0.01	-0.01	-0.003	-0.001	-0.0003	-0.0001
		1/N	**0.63**	**2.41**	**5.48**	**3.38**	**2.81**	**5.83**	**2.20**	**2.06**	**5.48**
	50	Opt	-52.75	-14.14	-4.34	-0.01	-0.01	-0.003	-0.001	-0.001	-0.0002
		1/N	-39.18	-2.37	**2.01**	**0.46**	**2.98**	**4.27**	**0.15**	**2.24**	**3.90**

Note: Difference in expected utility between the decision with judgment and the judgmental decision. A positive difference indicates that the decision with judgment outperforms the judgmental decision. The DGP underlying the Monte Carlo exercise is the same as that of Tables 1-3.

A positive difference indicates that the decision with judgment offers an improvement over the judgmental decision, by providing a higher expected utility on average, while a negative difference suggests the opposite. The results of the exercise are consistent with those of the previous subsection. When the judgmental decision coincides with the optimal allocation, the difference in expected utilities cannot be positive, as improving over the judgmental decision is impossible. In this scenario, the loss in utility associated with the maximum likelihood allocation ($\alpha = 1$) is significantly higher than when $\alpha < 1$. The results are striking when the risk aversion parameter is equal to 1 ($\gamma = 1$) and the portfolio contains a large number of assets ($N = 50$). Unlike the standard plug-in, maximum likelihood approach, the decision with judgment offers protection against Type I errors. The lower the level of significance, the greater the protection. However, this protection comes with a trade-off: reduced power of the test. When the judgmental decision is the

Table 5: Expected utility with rolling windows

			α=1			α=0.1			α=0.01		
	# assets	ã	180	300	600	180	300	600	180	300	600
γ=1	1	Opt	-2.87	-1.74	-0.81	-0.02	-0.02	-0.03	-0.0001	-0.0002	-0.002
		1/N	-1.89	-0.69	**0.09**	**0.14**	**0.19**	**0.28**	**0.01**	**0.02**	**0.05**
	5	Opt	-15.42	-8.89	-4.37	-0.11	-0.17	-0.34	-0.01	-0.01	-0.04
		1/N	-15.16	-8.65	-4.06	-0.09	-0.15	-0.27	-0.005	-0.01	-0.02
	20	Opt	-81.41	-41.89	-18.82	-0.52	-1.07	-2.73	-0.09	-0.25	-0.96
		1/N	-79.40	-40.28	-17.15	-0.42	-0.87	-2.31	-0.04	-0.14	-0.62
	50	Opt	-394.29	-151.49	-55.31	-1.50	-3.53	-9.91	-0.41	-1.40	-5.77
		1/N	-383.75	-139.87	-46.34	-0.73	-2.07	-6.24	**0.23**	-0.19	-2.38
γ=5	1	Opt	-0.57	-0.35	-0.17	-0.005	-0.004	-0.01	0	0	0
		1/N	**6.40**	**6.65**	**6.82**	**6.48**	**6.63**	**6.73**	**5.41**	**6.19**	**6.31**
	5	Opt	-3.08	-1.75	-0.87	-0.02	-0.03	-0.07	0.00	0.00	-0.01
		1/N	-0.82	**0.42**	**1.35**	**0.95**	**1.22**	**1.54**	**0.61**	**0.90**	**1.34**
	20	Opt	-16.28	-8.30	-3.78	-0.10	-0.21	-0.54	-0.02	-0.05	-0.19
		1/N	-14.44	-6.64	-1.97	**0.13**	**0.18**	**0.24**	**0.15**	**0.28**	**0.48**
	50	Opt	-77.90	-29.87	-11.18	-0.29	-0.71	-1.97	-0.08	-0.29	-1.15
		1/N	-73.83	-25.87	-6.95	**0.03**	**0.01**	-0.15	**0.23**	**0.36**	**0.56**
γ=10	1	Opt	-0.29	-0.17	-0.08	-0.002	-0.002	-0.004	0	0	0
		1/N	**14.20**	**14.34**	**14.42**	**14.28**	**14.34**	**14.37**	**14.14**	**14.16**	**14.18**
	5	Opt	-1.56	-0.88	-0.43	-0.01	-0.02	-0.04	0.00	0.00	0.00
		1/N	**3.50**	**4.16**	**4.62**	**3.92**	**4.30**	**4.62**	**3.76**	**4.16**	**4.47**
	20	Opt	-8.14	-4.16	-1.87	-0.05	-0.10	-0.27	-0.01	-0.02	-0.09
		1/N	-3.56	**0.37**	**2.64**	**1.64**	**2.37**	**3.17**	**1.61**	**2.34**	**3.16**
	50	Opt	-39.23	-15.12	-5.64	-0.15	-0.35	-1.00	-0.04	-0.14	-0.58
		1/N	-31.12	-6.61	**2.91**	**1.46**	**2.85**	**4.78**	**1.53**	**3.03**	**5.00**

Note: The exercise mirrors to that of Table 4, but it now employs rolling windows to estimate the judgmental decision.

equal weight portfolio and it is significantly worse than the optimal allocation, the decision with judgment at $\alpha = 10\%$ tends to outperform that at $\alpha = 1\%$.

Table 5 presents a similar exercise to that of Table 4, with the key difference being the use of rolling windows of estimation. We simulate a time series of 1,000 returns and assess the performance by using windows of 180, 300 and 600 observations to estimate the decision with judgment. After computing the expected utility, the estimation window is rolled over by one observation to re-estimate the decision with judgment. The process is repeated until the end of the sample is reached. This entire procedure is conducted 1,000 times, and the overall average utilities are computed for both the decision with judgment and the judgmental decision. The results are qualitatively similar to those of Table 4, but with an intriguing twist: expanding the estimation window does not necessarily enhance the performance of the decision with judgment. In fact, when the judgmental allocation is optimal, increasing the estimation window leads to substantially worse performance. This

Table 6: Data sets used in the pseudo out-of-sample exercise

Dataset		T	N
1	25_Portfolios_5x5_1Factor	702	21
2	25_Portfolios_5x5_MOM	706	24
3	F-F_Research_Data_Factors	941	3
4	Industry	941	11
5	International	379	9
6	SPSectors	264	11

Note: Datasets utilized in the pseudo out-of-sample exercise. The number of observations is denoted by T and the number of assets included in each portfolio by N.

is due to the persistence introduced by the overlapping rolling windows: when the test incorrectly rejects the null hypothesis that the judgmental decision is optimal, it is likely to continue rejecting it as the window shifts forward by one observation.

The results from this exercise are important as they represent a more realistic benchmark for evaluating pseudo out-of-sample exercises performed with real data. We next turn such an exercise.

6 Pseudo out-of-sample performance

We consider similar datasets to those used in DeMiguel et al. (2009), as reported in Table 6.

The portfolios are the same as in DeMiguel et al. (2009), which in turn are taken from Ken French's website. The 25 portfolios are the portfolios sorted by size and book-to-market, excluding the five portfolios with the largest firms, as they highly correlated with the factors.

We conduct the same out-of-sample exercise as in Table 5, considering three estimation windows of 180, 300 and 600 observations, along with three levels of risk aversion, $\gamma = 1, 5, 10$. Since we are now dealing with real data, the optimal allocation is unknown. To address this, we construct a pseudo-optimal allocation by estimating the portfolio weights using all available observations in the sample. The alternative judgmental decision considered is the equally weighted portfolio.

The decisions with judgment are estimated using the observations within the estimation window. We next compute the expected utility associated with these decisions by using the full sample mean and variance-covariance matrix, which are used to construct the pseudo optimal decision. This process is repeated for the judgmental decisions as well. The window is subsequently rolled over by one observation, and the whole procedure is repeated until the penultimate observation is reached. Finally, we average the expected utility over the entire out-of-sample period and report the differences between the expected utilities of the decisions with judgment and the judgmental decision. As in the previous section, a positive difference indicates that the decision with judgment outperforms the judgmental decision, while a negative difference suggests the opposite.

The results, presented in Table 7, are broadly consistent with those observed in the Monte Carlo simulation. First, the pseudo optimal decision cannot be improved by construction. However, the standard plug-in maximum likelihood estimator's underperformance is significantly more pronounced than that of the decision with judgment. Decisions with judgment, when using a level of significance less than 1, protect the decision-maker against Type I errors.

Second, aligning with the findings of DeMiguel et al. (2009), the maximum likelihood decision generally underperforms compared to the equal weight portfolio. It is only when the risk aversion is high ($\gamma = 10$) that the maximum likelihood allocation does better.

Third, the decision with judgment almost always outperforms the equal weight portfolio. Furthermore, when the judgmental decision is optimal, the losses in terms of expected utility are far more limited than those for the plug-in allocation. An exception occurs for dataset #2, which includes an outlier that skews the portfolio allocation's performance. This outlier results in atypical portfolio returns that disproportionately affect the expected utility calculations. While the decision with judgment typically mitigates Type I errors, it might still be affected by extreme observations. As a result, its relative advantage over the equal weight portfolio may be temporarily reduced. Addressing this anomaly requires additional robustness checks or adjustments to effectively account for such outliers.

Table 7: Pseudo out-of-sample performance expected utility

			α=1			α=0.1			α=0.01		
	Dataset	ã	180	300	600	180	300	600	180	300	600
γ=1	1	Opt	-302	-73	-7	-159	-31	-1	-105	-16	0
		1/N	-244	-15	**51**	-116	**14**	**53**	-70	**23**	**51**
	2	Opt	-15772	-6344	-2382	-10653	-4459	-1805	-8561	-3641	-1540
		1/N	-15414	-5986	-2025	-10639	-4305	-1535	-8693	-3579	-1310
	3	Opt	-25	-11	-4	-5	-1	0	-1	0	0
		1/N	-21	-7	-1	-4	**1**	**2**	0	**1**	**2**
	4	Opt	-323	-150	-48	-121	-54	-15	-54	-22	-5
		1/N	-311	-138	-36	-116	-49	-7	-52	-19	**2**
	5	Opt	-19	NA	NA	-2	NA	NA	0	NA	NA
		1/N	-6	NA	NA	**5**	NA	NA	**3**	NA	NA
	6	Opt	-41	NA	NA	-6	NA	NA	0	NA	NA
		1/N	-5	NA	NA	**18**	NA	NA	**19**	NA	NA
γ=5	1	Opt	-60	-15	-2	-32	-6	0	-21	-3	0
		1/N	-47	-1	**12**	-22	**5**	**13**	-13	**6**	**12**
	2	Opt	-3154	-1269	-477	-2131	-892	-361	-1712	-728	-308
		1/N	-3082	-1196	-404	-2130	-863	-307	-1742	-719	-262
	3	Opt	-5	-2	-1	-1	-0.3	-0.1	-0.3	-0.1	-0.1
		1/N	-3	-1	**1**	-0.3	**0.4**	**1**	**0.1**	**0.2**	**1**
	4	Opt	-65	-30	-10	-24	-11	-3	-11	-5	-1
		1/N	-59	-24	-4	-21	-7	**1**	-8	-2	**2**
	5	Opt	-4	NA	NA	-0.4	NA	NA	-1	NA	NA
		1/N	**2**	NA	NA	**3**	NA	NA	**2**	NA	NA
	6	Opt	-8	NA	NA	-0.1	NA	NA	-0.1	NA	NA
		1/N	**1**	NA	NA	**5**	NA	NA	**4**	NA	NA
γ=10	1	Opt	-30	-7	-1	-16	-3	0	-10	-2	0
		1/N	-16	**7**	**13**	-6	**9**	**13**	-2	**9**	**13**
	2	Opt	-1577	-634	-238	-1065	-446	-181	-856	-364	-154
		1/N	-1536	-593	-197	-1069	-433	-150	-877	-363	-128
	3	Opt	-2	-1	-0.4	-0.5	-0.1	-0.01	-0.1	-0.02	0
		1/N	**1**	**2**	**3**	**2**	**2**	**3**	**1**	**1**	**2**
	4	Opt	-32	-15	-5	-12	-5	-1	-5	-2	0
		1/N	-21	-4	**7**	-5	**3**	**8**	**0.1**	**5**	**8**
	5	Opt	-2	NA	NA	-0.2	NA	NA	-0.02	NA	NA
		1/N	**7**	NA	NA	**7**	NA	NA	**6**	NA	NA
	6	Opt	-4	NA	NA	-1	NA	NA	-0.04	NA	NA
		1/N	**6**	NA	NA	**7**	NA	NA	**6**	NA	NA

Note: Difference between out-of-sample expected utilities of the decision with judgment and of the judgmental decision.

7 Conclusion

The explicit formulation of judgment in terms of a benchmark asset allocation can be used to address the empirical challenges that often arise with traditional econometric methods. Utilizing the statistical decision theoretical framework developed by Manganelli (2025), the decision with judgment first tests whether the benchmark allocation is optimal. This is accomplished by testing whether the gradient of the expected utility evaluated at the benchmark allocation is equal to zero. In case of rejection, the decision with judgment selects the asset allocation which coincides with the boundary of the confidence interval of the gradient.

The core innovation of this approach is the use of judgmental decisions, which are benchmark allocations that decision-makers aim to outperform. By formulating the problem within a hypothesis testing framework, we can utilize maximum likelihood estimates to evaluate whether these judgmental decisions are indeed optimal. The level of significance provides a systematic way to control Type I errors, ensuring that deviations from the benchmark are statistically justified. Stopping at the boundary of the confidence interval of the gradient protects the decision-maker against estimation error, providing a significant advantage over traditional plug-in estimators.

Our Monte Carlo simulations revealed that this method is robust across various scenarios, including different number of assets in the portfolio and degrees of risk aversion. The simulations showed that the asymptotic approximation of the test statistic is accurate in size. Its power depends on the accuracy of the benchmark allocation (which in turn is linked to the degree of risk aversion of the decision-maker) and the number of assets included in the portfolio. Moreover, it consistently outperforms the $1/N$ benchmark allocation, particularly in scenarios where the judgmental decision is suboptimal.

The pseudo out-of-sample exercises using real datasets further validated these findings. The decision with judgment demonstrated superior performance compared to both the traditional plug-in estimators and equally weighted allocations. This suggests that incor-

porating judgment into asset allocation not only enhances theoretical soundness but also offers practical benefits in real-world applications.

While the results are promising, future research could explore the extension of this framework to more complex financial models (such as time-varying variances) and different market conditions. The incorporation of judgment into asset allocation provides a compelling alternative to traditional methods. It offers a rigorous yet flexible framework that enhances decision-making under uncertainty, potentially leading to improved investment outcomes.

References

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

Brandt, M. (2009). Portfolio choice problems. In Y. Ait-Sahalia and L. P. Hansen (Eds.), *Handbook of Financial Econometrics*. North Holland.

DeMiguel, V., L. Garlappi, and R. Uppal (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy. *The Review of Financial Studies 22*(5), 1915–1953.

Ellsberg, D. (1961). Risk, ambiguity and the Savage axioms. *The Quarterly Journal of Economics 75*(4), 643–669.

Epstein, L. and M. Schneider (2010). Ambiguity and asset markets. *Annual Review of Financial and Economics 2*, 315–346.

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

Gilboa, I. and M. Marinacci (2013). Ambiguity and the Bayesian paradigm. In D. Acemoglu, M. Arellano, and E. Dekel (Eds.), *Advances in Economics and Econometrics: Theory and Applications*. 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.

Ilut, C. and M. Schneider (2023). Chapter 24 - Ambiguity. In R. Bachmann, G. Topa, and W. van der Klaauw (Eds.), *Handbook of Economic Expectations*, pp. 749–777. Academic Press.

Jobson, J. and B. Korkie (1980). Estimation for markowitz efficient portfolios. *Journal of the American Statistical Association 75*(371), 544–554.

Lütkepohl, H. (2005). *The New Introduction to Multiple Time Series Analysis*. Springer-Verlag Berlin.

MacKinlay, A. C. and L. Pastor (2000). Asset pricing models: Implications for expected returns and portfolio selection. *The Review of Financial Studies 13*(4), 883–916.

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

Manganelli, S. (2023). Double conditioning: the hiddent connection between bayesian and classical statistics. *European Central Bank 2786*, 1–24.

Manganelli, S. (2025). Statistical decision functions with judgment. *Journal of Economic Theory 223*, 105940.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance 39*, 47–61.

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

Sharpe, W. (1964). Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance 19*(3), 425–442.

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