Forecasting With Judgment

Simone Manganelli
DG-Research, European Central Bank, Kaiserstrasse 29, D-60311 Frankfurt am Main, Germany
(*simone.manganelli@ecb.int*)

This article shows how to account for nonsample information in the classical forecasting framework. We explicitly incorporate two elements: a default decision and a probability reflecting the confidence associated with it. Starting from the default decision, the new estimator increases the objective function only as long as its first derivatives are statistically different from zero. It includes as a special case the classical estimator and has clear analogies with Bayesian estimators. The properties of the new estimator are studied with a detailed risk analysis. Finally, we illustrate its performance with applications to mean-variance portfolio selection and to GDP forecast.

KEY WORDS: Asset allocation; Decision under uncertainty; Estimation; Overfitting.

1. INTRODUCTION

Forecasting is intrinsically intertwined with decision making. Forecasts help agents make decisions when facing uncertainty. Forecast errors impose costs on the decision maker, to the extent that different forecasts command different decisions. It is generally assumed that agents wish to minimize the expected cost associated with these errors (see chapter VI of Haavelmo 1944; Granger and Newbold 1986; Granger and Machina 2006). Classical forecasts are obtained as the minimizers of the sample equivalent of the unobservable expected cost.

This theory neglects one important element of decision problems in practice. It does not *explicitly* account for the nonsample information available to the decision maker, even though subjective judgment often plays an important role in many real world forecasting processes. In this article we show how to incorporate the nonsample information in the classical econometric analysis.

We explicitly account for the nonsample information by introducing two elements in the estimation framework. The first element is a default decision for the decision problem underlying the forecasting exercise. A default decision is the choice of the decision maker who does not have access to statistics. It may be the result of experience, introspection, habit, herding, or just pure guessing. It is actually how most people make decisions in their daily life. The second element is a probability reflecting the confidence of the decision maker in this default decision, in the sense that it determines the amount of statistical evidence needed to discard the default decision. These elements serve to summarize the nonsample information available to the decision maker, and allow us to formalize the interaction between judgment and data in the forecasting process.

The estimator is then constructed by first testing whether the default parameters (i.e., the model parameters implied by the default decision) are significantly different from those that maximize the in-sample objective function. This is equivalent to testing whether its first derivatives evaluated at the default parameters are statistically equal to zero at the given confidence level. If this is the case, the default decision is chosen. Otherwise, the objective function is increased as long as its first derivatives are significantly different from zero, and a new, model-based decision is obtained. The resulting estimator—which will be called "subjective classical estimator"—will be on the perimeter of the confidence set, unless the default parameters are within the confidence set, in which case the estimator equals the default parameters.

Classical estimators set the first derivatives equal to zero. They can, therefore, be obtained as a special case of our estimator by choosing a confidence level equal to zero, which corresponds to ignoring the default decision. Moreover, under standard regularity conditions the new estimator is shown to be consistent. As the sample size grows, the true objective function is approximated with greater precision, and the default decision becomes less and less relevant.

Default parameters and the associated confidence level are routinely used in classical econometrics to test hypotheses about specific parameter values of a model. They are used, however, in a fundamentally different way with respect to the procedure suggested in this article. The classical procedure first derives the statistical properties of the estimator and then tests whether the estimate is different from the default parameters, for the prespecified confidence level. The problem with this procedure is that, in case of rejection of the null hypothesis, it does not specify how to choose the parameters within the nonrejection region. In the procedure proposed in this article, the null hypothesis is that the first derivatives of the objective function evaluated at the default parameters are equal to zero, while the alternative hypothesis is that they are not. If the null is rejected, the alternative tells that at the default parameters it is possible to increase the expected utility. This holds true only until the first derivatives stop being statistically different from zero.

Pretest estimators clearly illustrate the problems associated with the classical procedure. They test whether the default parameters are statistically different from the classical estimator, and in case of rejection revert to the classical estimator. A fundamental problem with this approach is that rejection of the null hypothesis only signals that the first derivatives evaluated at the default parameters are statistically different from zero, and, therefore, that the objective function can be significantly improved. However, the objective function can be improved *only*

Journal of Business & Economic Statistics
October 2009, Vol. 27, No. 4
DOI: 10.1198/jbes.2009.08052

up to the point where the first derivatives are no longer significantly different from zero. Beyond that point, the null hypothesis that the objective function is not increasing cannot be rejected any longer at the chosen confidence level.

We illustrate this insight with a detailed risk analysis of the proposed estimator, comparing its performance to the classical and pretest estimators. We show how our estimator, unlike the classical and pretest estimators, outperforms any given default decision in a precise statistical sense.

An important issue regards the relationship between this theory and Bayesian econometrics. We argue that our theory and Bayesian techniques offer two alternative methods to incorporate nonsample information in the econometric analysis. We show how in two special cases (when there is no information besides the sample and when there is certainty about model parameters) the two estimators coincide. In intermediate cases, the choice depends on how the nonsample information is formalized. Bayesian econometrics should be used whenever the nonsample information takes the form of a prior probability distribution. In contrast, the estimator proposed in this article can be used whenever the nonsample information is expressed in terms of a default decision and a confidence associated to it. From this perspective, the choice between Bayesian and classical econometrics is not an issue to be settled by econometricians, but rather by the decision maker through the format in which she/he provides the nonsample information.

A potential problem with Bayesian econometrics is that formulation of priors is not always obvious and may in some cases put a heavy burden on the decision maker (see Durlauf 2002 for a discussion of the difficulties in eliciting prior distributions). We illustrate with two examples how our estimator may be relatively straightforward to implement. In the first example, we show how the forecasting framework proposed in this article can be used to tackle some of the well-known implementation problems of mean-variance portfolio selection models. For a given benchmark portfolio (the default decision, in the terminology used before), we derive the associated optimal portfolio which increases the empirical expected utility as long as the first derivatives are statistically different from zero. We show with an out of sample exercise how the proposed estimator statistically outperforms the given benchmark. Consistently with the insight from the risk analysis mentioned before, this is not always the case for the classical mean-variance optimizers.

In the second example, we provide an application to GDP growth rate forecasts. Econometric models are complicated functions of parameters which are often devoid of economic meaning. It may, therefore, be difficult to express a default decision in terms of these parameters. We suggest a simple and intuitive strategy to map the default forecast on the variable of interest to the decision maker into values for the parameters of the econometrician's favorite model. Specifically, these "judgmental parameter values" are obtained by maximizing the objective function subject to the constraint that the forecast implied by the model is equal to the default forecast. We illustrate how this works in the context of a simple autoregressive model.

The article is structured as follows. In the next section, we use a stylized statistical model to highlight the problems associated to classical estimators. In Section 3, we build on this stylized model to develop the heuristics behind the new estimator. Section 4 presents the risk analysis of the proposed estimator. Section 5 contains a formal development of the new theory and discusses its relationship with Bayesian econometrics. The empirical applications are in Section 6. Section 7 concludes.

2. THE PROBLEM

Classical estimators approximate the expected utility function with its sample equivalent. While asymptotically this approximation is perfect, in finite samples it is not. The quality of the finite sample approximation—which is out of the econometrician's control—will crucially determine the quality of the forecasts.

Assume that $\{y_t\}_{t=1}^T$ is a series of iid normally distributed observations with unknown mean θ^0 and known variance equal to 1. We are interested in the forecast θ of y_{T+1}, using the information available up to time T. Let's denote the forecast error by $e \equiv y_{T+1} - \theta$. Suppose that the agent quantifies the (dis)utility of the error with a quadratic utility function, $U(e) \equiv -e^2$. The optimal forecast maximizes the expected utility:

$$\max_{\theta} E[-(y_{T+1} - \theta)^2]. \tag{1}$$

Setting the first derivative equal to zero, the optimal forecast is given by the expected value of y_{T+1}, leading to the classical estimator $\hat{\theta}_T \equiv T^{-1}\sum_{t=1}^T y_t$. But $\hat{\theta}_T$ is the maximizer of $T^{-1}\sum_{t=1}^T[-(y_t - \theta)^2]$ and the problem can be rewritten as:

$$\max_{\theta}\{E[-(y_{T+1} - \theta)^2] + \varepsilon_T(\theta)\}, \tag{2}$$

where $\varepsilon_T(\theta) \equiv E[(y_{T+1} - \theta)^2] - T^{-1}\sum_{t=1}^T[(y_t - \theta)^2]$. $\varepsilon_T(\theta)$ is the error induced by the finite sample approximation of the expected utility function, which by the Law of Large Numbers converges to zero only as T goes to infinity. Therefore, in finite samples, classical estimators do not maximize the expected utility, but also an error term $\varepsilon_T(\theta)$ which vanishes only asymptotically.

3. AN ALTERNATIVE FORECASTING STRATEGY

Assume now that the decision maker has nonsample information that can be expressed in terms of a default decision $\tilde{\theta}$ and a probability α summarizing the confidence in this decision.

The first order condition of the optimal forecast problem in Equation (1) is:

$$E[y_{T+1} - \theta] = 0. \tag{3}$$

The sample equivalent of this expectation evaluated at $\tilde{\theta}$ is:

$$f_T(\tilde{\theta}) \equiv \hat{y}_T - \tilde{\theta}, \tag{4}$$

where $\hat{y}_T \equiv T^{-1}\sum_{t=1}^T y_t$. $f_T(\tilde{\theta})$ is the sample mean of the first derivatives of the expected utility function. It is a random variable which may be different from zero just because of statistical error. Under the null hypothesis that $\tilde{\theta}$ is the optimal estimator, $f_T(\tilde{\theta}) \sim \mathrm{N}(0, 1/T)$.

Given the default decision $\tilde{\theta}$ and confidence level α, it is possible to test the null hypothesis $H_0: \tilde{\theta} = \theta^0$, which is equivalent to $H_0: E[f_T(\tilde{\theta})] = 0$. If the null cannot be rejected, there is not enough statistical evidence against $\tilde{\theta}$ and $\tilde{\theta}$ becomes the adopted decision. On the other hand, rejection of the null signals that the first derivative is significantly different from zero and, therefore, that the objective function can be increased. This, however, is true only up to the point where the first derivative stops being significantly different from zero, i.e., at the point θ_T^* where $f_T(\theta_T^*)$ is exactly equal to its critical value.

We can formalize the above discussion as follows. For a given confidence level α, let $\pm\kappa_{\alpha/2}$ denote the corresponding standard normal critical values and $\pm\eta_{\alpha/2}(T) \equiv \pm\sqrt{T^{-1}}\kappa_{\alpha/2}$. The resulting estimator, which we refer to as *subjective classical estimator*, is:

$$\theta_T^* = \begin{cases} \hat{y}_T - \eta_{\alpha/2}(T) & \text{if } \hat{y}_T - \tilde{\theta} > \eta_{\alpha/2}(T) \\ \tilde{\theta} & \text{if } |\hat{y}_T - \tilde{\theta}| < \eta_{\alpha/2}(T) \\ \hat{y}_T + \eta_{\alpha/2}(T) & \text{if } \hat{y}_T - \tilde{\theta} < -\eta_{\alpha/2}(T). \end{cases} \quad (5)$$

Note that $\eta_{\alpha/2}(T)$ converges to zero as T goes to infinity. Therefore, θ_T^* converges to θ^0 in probability.

Remark 1 (Economic interpretation). This estimator has a natural economic interpretation in terms of the expected cost/utility function used in the forecasting problem. For a given default decision $\tilde{\theta}$ and confidence level α, it answers the following question: Can the forecaster increase his/her expected utility in a statistically significant way? If the answer is no, i.e., if one cannot reject the null that the first derivative evaluated at $\tilde{\theta}$ is equal to zero, $\tilde{\theta}$ should be taken as the forecast. If, on the contrary, the answer is yes, the econometrician will move the parameter θ as long as the first derivative is statistically different from zero. She/he will stop only when θ_T^* is such that the empirical expected utility cannot be increased any more in a statistically significant way. This happens exactly at the boundary of the confidence interval.

Remark 2 (Nonsample information). Both $\tilde{\theta}$ and α are exogenous to the statistical problem. They summarize the nonsample information available to the decision maker, and represent subjective elements in the analysis. $\tilde{\theta}$ represents the choice of the decision maker who does not have access to statistics. It may be the result of experience, introspection, habit, herding, or just pure guessing. It is actually how most people make decisions in their daily life. The confidence level α reflects the confidence of the forecaster in the default decision and should be inversely related to the knowledge of the environment in which the forecast takes place: the better the knowledge of such environment, the higher the confidence in the default decision, the lower α. The confidence level α determines the amount of statistical evidence needed to reject the default decision. In the jargon of hypothesis testing, it reflects the willingness of committing Type I errors, i.e., of rejecting the null when $\tilde{\theta}$ is indeed the optimal forecast. One should be careful in not interpreting α as the probability that the correct decision θ^0 is actually equal to the default decision $\tilde{\theta}$, as with continuous random variables this would be a meaningless probabilistic statement.

Note that in the classical paradigm there is no place for default decisions and therefore $\alpha = 1$: in this case $\kappa_{\alpha/2} = 0$ and θ_T^* is simply the solution obtained by setting the first derivative of Equation (4) equal to zero.

Remark 3 (Relationship with pretest estimators). Pretest estimators would first test the null hypothesis $H_0: \theta^0 = \tilde{\theta}$, and in case of rejection revert to the classical estimator. The choice is, therefore, either $\tilde{\theta}$ or $\hat{\theta}_T$. However, rejection of the null hypothesis just signals that the first derivative at $\tilde{\theta}$ is significantly different from zero, and, therefore, that the objective function can be increased in a statistically significant way. Note that $\eta_{\alpha/2}(T)$ is the critical point beyond which all the null hypotheses with $E[f_T(\theta)] \leq 0$ are rejected at the chosen confidence level. Similarly, $-\eta_{\alpha/2}(T)$ is the critical point below which all the null hypotheses with $E[f_T(\theta)] \geq 0$ are rejected. For any $\bar{\theta}$ such that $-\eta_{\alpha/2}(T) < f_T(\bar{\theta}) < \eta_{\alpha/2}(T)$, instead, neither $H_0: E[f_T(\bar{\theta})] \leq 0$ nor $H_0: E[f_T(\bar{\theta})] \geq 0$ can be rejected. In this case there is not enough statistical evidence guaranteeing that by moving closer to $\hat{\theta}_T$ the objective function in population is increased. This suggests an alternative interpretation of estimator in Equation (5). It is the parameter value that is closest to the decision maker's default parameter, subject to the constraint that the mean of the first derivative of the objective function at the parameter value is not significantly different from zero.

Remark 4 (Relationship with the Burr estimator). In the special example considered in this section, the estimator found in Equation (5) coincides with the Burr estimator, as discussed by Magnus (2002). Magnus arrived at this estimator following a completely different logic. (Hansen 2007 also arrives at the same estimator, by shrinking the parameters towards some gravity point, until the criterion function is reduced by a prespecified amount.) He shows that the Burr estimator is the minimax regret estimator of a large class of estimators, defined by the class of distribution functions $\varphi(x;\beta,\gamma) = 1 - (1 + (x^2/c^2)^\beta)^{-\gamma}$, where $\beta > 0$, $\gamma > 0$, and c is a scale parameter, which corresponds to $\eta_{\alpha/2}(T)$ in Equation (5). A key difference between the two approaches is that while in Magnus' case c is a free parameter which should be optimized by the econometrician, in our approach c is exogenous to the statistical problem, as it is provided by the decision maker (see the discussion in Remark 2). As discussed by Magnus (2002), the estimator in Equation (5) is kinked and, therefore, inadmissible (see theorem A.6 in Magnus 2002). Finding an admissible correction is not trivial. However, it is often the case that the improvement provided by such corrections is negligible.

Remark 5 (Relationship with forecast combination). The estimator in Equation (5) can be seen as an alternative way to combine the two different forecasts $\hat{\theta}_T$ and $\tilde{\theta}$ (see, for instance, Granger and Newbold 1986, chapter 9). An alternative forecast combination which does not rely on the availability of out of sample forecasts is to define $\hat{\theta}_T^{**} = \frac{\sigma_T^2}{\gamma^2+\sigma_T^2}\tilde{\theta} + \frac{\gamma^2}{\gamma^2+\sigma_T^2}\hat{\theta}_T$, where $\sigma_T^2 \equiv \hat{V}[\hat{\theta}_T]$ and γ^2 denotes the variance of $\tilde{\theta}$. This estimator is also consistent for θ^0 if $\gamma^2 > 0$, as $\sigma_T^2 \to 0$ as $T \to \infty$, and it also incorporates the decision maker's default parameter and a confidence level for this. (There is a precise relationship between α and γ^2, since $\alpha \to 0 \Longrightarrow \gamma^2 \to 0$, $\alpha \to 1 \Longrightarrow \gamma^2 \to \infty$.) Gonzalez, Hubrich, and Teräsvirta (2009) proposed yet another method to incorporate nonsample information based on a penalized likelihood. A key practical difference between these approaches is that the estimator proposed in this article uses information from the classical estimate only when the default decision is significantly far away from the optimum. The other estimators always uses a bit of the default decision and a bit of the classical estimate.

4. RISK ANALYSIS

We evaluate the performance of the estimator proposed in the previous section by comparing its risk function with that of some popular alternatives.

Let y_1 be drawn from a univariate normal distribution with unknown mean θ^0 and variance 1. The goal is to estimate the mean θ^0 given the single observation y_1. It corresponds to the problem discussed in Section 2 with $T = 1$, and has been treated extensively by Magnus (2002). Despite its appearance, solutions to this problem have important practical implications. The problem was shown by Magnus and Durbin (1999) to be equivalent to the problem of estimating the coefficients of a set of explanatory variables in a linear regression model, when there is doubt whether additional regressors should be included in the model.

We compare the risk properties of the estimator proposed in Section 3 with the standard OLS estimator and the pretest estimator. Specifically, we compare the estimator θ_1^* proposed in Equation (5), referred to as *subjective classical estimator*, with the following *pretest estimator*:

$$\hat{\theta}_1^P = \begin{cases} \tilde{\theta} & \text{if } |y_1 - \tilde{\theta}| < \kappa_{\alpha/2} \\ y_1 & \text{if } |y_1 - \tilde{\theta}| > \kappa_{\alpha/2}. \end{cases}$$

Note that by setting $\alpha = 1$ and $\alpha = 0$ in both the subjective classical and pretest estimators, we get as special cases the OLS estimator and the default decision, respectively.

In the case of quadratic loss function, the risk associated to an estimator $f(y)$ is defined as:

$$R(\theta^0; f(y)) = E_{\theta^0}\left[(f(y) - \theta^0)^2\right].$$

In Figure 1 we report the risks associated to the pretest and subjective classical estimators, with $\tilde{\theta} = 0$ and $\alpha = 1$, 0.10, and 0. We notice that for values of θ^0 sufficiently close to $\tilde{\theta}$, both OLS and pretest estimators have higher risk than the subjective classical estimator with $\alpha = 0.10$. This holds also for values of θ^0 around 1, the value at which the risks of the OLS estimator and the default decision coincide. Furthermore, the new estimator has bounded risk (unless $\alpha = 0$, which corresponds to the default decision) and has always lower risk than the default decision, except for values of $\tilde{\theta}$ very close to θ^0. (In the example shown in Figure 1 the upper bound for the subjective classical estimator is approximately 3.7.) This is a very appealing feature of our estimator, as it guarantees that it will outperform any given default decision in a precise statistical sense.

The decision maker can control via the confidence level α the degree of underperformance when $\tilde{\theta}$ is close to θ^0. The effect of choosing different values for α is to move the risk function of the subjective classical estimator towards its extremes. The higher the confidence in the default decision (i.e., the lower the α), the steeper its risk function, converging to the risk of the default decision as $\alpha \to 0$. The lower the confidence (i.e., the higher the α), the flatter its risk function, converging to the risk of OLS as $\alpha \to 1$.

The lower the α, the better the performance of the new estimator for values of θ^0 close to $\tilde{\theta}$, and the worse for values of θ^0 farther away from $\tilde{\theta}$. This implies that high confidence in the default decision $\tilde{\theta}$ pays off if its value is close enough to θ^0, but comes at the cost of higher risk whenever $\tilde{\theta}$ is far away from θ^0.

Notice that in the previous example a default decision based on the OLS estimator would never be rejected by the data (since OLS sets the first derivatives equal to zero), but it would also have no value added. High confidence in a bad default decision, on the other hand, would inevitably result in poor forecasts (in small samples). Therefore, formulating a good default decision may be as important as having a good econometric model.

5. INCORPORATING JUDGMENT INTO CLASSICAL ESTIMATION

In this section we generalize the analysis of the previous sections. We formally define a new estimator which depends on

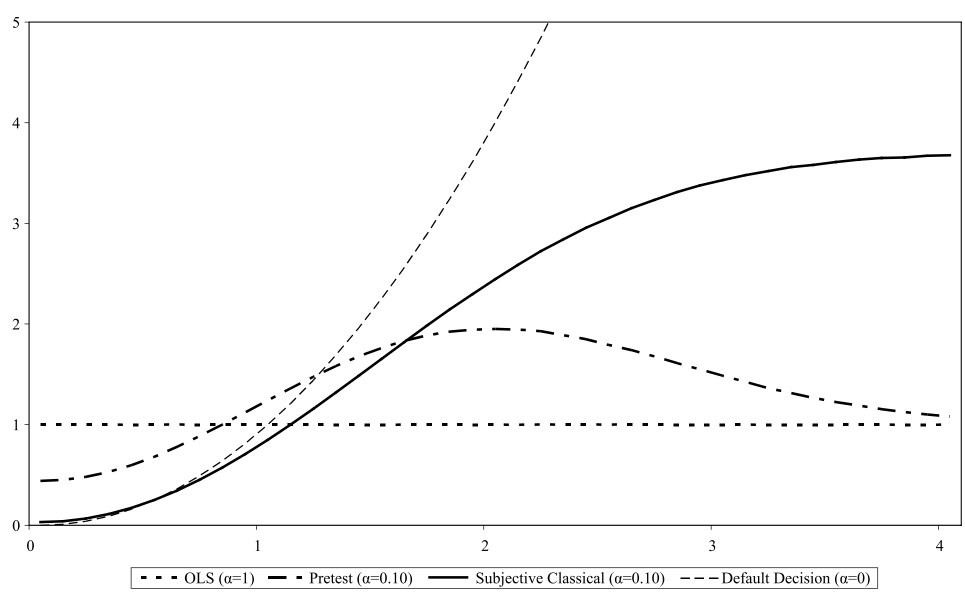

Figure 1. Risk comparison of different estimators.

a default decision and the confidence associated to it, and establish its relationships with classical estimators. This new estimator is obtained by adding a constraint on the first derivatives to the classical optimization problem. We also discuss the relationship with Bayesian econometrics.

As argued in the Introduction, optimal forecasts maximize the expected utility of the decision maker, which depends on the actions to be taken. Following the framework of Newey and McFadden (1994), denote with $\hat{U}_T(\boldsymbol{\theta})$ the finite sample approximation of the expected utility, which depends on the decision variables $\boldsymbol{\theta}$, data and sample size. $\boldsymbol{\theta}$ is a vector belonging to the k-dimensional parameter space Θ. We assume the following:

Condition 1 (Uniform convergence). $\hat{U}_T(\boldsymbol{\theta})$ converges uniformly on Θ in probability to $U_0(\boldsymbol{\theta})$, that is $\sup_{\boldsymbol{\theta}\in\Theta} |\hat{U}_T(\boldsymbol{\theta}) - U_0(\boldsymbol{\theta})| \xrightarrow{p} 0$.

Condition 2 (Identification). $U_0(\boldsymbol{\theta})$ is uniquely maximized at $\boldsymbol{\theta}^0$.

Condition 3 (Compactness). Θ is compact.

Condition 4 (Continuity). $U_0(\boldsymbol{\theta})$ is continuous.

These are the standard conditions needed for consistency results of extremum estimators (see theorem 2.1 of Newey and McFadden 1994). The classical estimator maximizes the empirical expected utility:

Definition 1 (Classical estimator). The *classical estimator* is $\hat{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta}} \hat{U}_T(\boldsymbol{\theta})$.

Before defining the new estimator, we impose the following conditions:

Condition 5. $\boldsymbol{\theta}^0 \in interior(\Theta)$.

Condition 6 (Differentiability). $\hat{U}_T(\boldsymbol{\theta})$ is continuously differentiable.

Condition 7 (Asymptotic normality). $\sqrt{T}\nabla_\theta \hat{U}_T(\boldsymbol{\theta}^0) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$.

Condition 8 (Convexity). $U_0(\boldsymbol{\theta})$ is globally convex in $\boldsymbol{\theta}$.

Convexity of the objective function in $\boldsymbol{\theta}$ is needed to ensure consistency, as otherwise the estimator could stop at a local maximum. More precisely convexity is needed because the test for optimality is based on the first order conditions. Since first derivatives describe only local behavior, without the convexity assumption setting the first derivative equal to zero is only a necessary, but not sufficient, condition for optimality. It may be possible to eliminate the convexity assumption by basing the test for optimality on a likelihood ratio test. However, reasoning in terms of first derivatives helps building intuition and motivating the proposed approach. The others are technical conditions typically imposed to derive asymptotic normality results.

Define the following convex linear combination, which shrinks from the default decision $\tilde{\boldsymbol{\theta}}$ to the classical estimate $\hat{\boldsymbol{\theta}}_T$:

$$\boldsymbol{\theta}^*(\lambda) \equiv \lambda\hat{\boldsymbol{\theta}}_T + (1-\lambda)\tilde{\boldsymbol{\theta}}, \qquad \lambda \in [0, 1]. \tag{6}$$

Denote with $\hat{\boldsymbol{\Sigma}}_T$ a consistent estimate of $\boldsymbol{\Sigma}$. Define [note that the test statistic in (7) relies on an asymptotic approximation to its distribution. If it is suspected that with the available sample size such approximation may be poor, one could resort to bootstrap methods to improve the accuracy of the estimator]:

$$\hat{z}_T(\boldsymbol{\theta}^*(\lambda)) \equiv T\nabla'_\theta \hat{U}_T(\boldsymbol{\theta}^*(\lambda))\hat{\boldsymbol{\Sigma}}_T^{-1}\nabla_\theta \hat{U}_T(\boldsymbol{\theta}^*(\lambda)). \tag{7}$$

Notice that given the realization of $\hat{\boldsymbol{\theta}}_T$, $\boldsymbol{\theta}^*(\lambda)$ for $\lambda \in [0, 1]$ represents a point on the segment joining $\hat{\boldsymbol{\theta}}_T$ and $\tilde{\boldsymbol{\theta}}$. We can, therefore, test whether any of these points is equal to $\boldsymbol{\theta}^0$. Under the null hypothesis $H_0 : \boldsymbol{\theta}^0 = \boldsymbol{\theta}^*(\lambda)$, we have:

$$\hat{z}_T(\boldsymbol{\theta}^*(\lambda)) = \hat{z}_T(\boldsymbol{\theta}^0) \xrightarrow{d} \chi^2_k.$$

The new estimator is defined as follows:

Definition 2 (Subjective Classical Estimator). Let $\tilde{\boldsymbol{\theta}}$ denote the default decision and $\alpha \in [0, 1]$ the confidence level associated to it. Define $\boldsymbol{\theta}^*(\lambda)$ and $\hat{z}_T(\boldsymbol{\theta}^*(\lambda))$ as in Equations (6) and (7), respectively, and let $\eta_{\alpha,k}$ denote the χ^2_k critical value associated to the confidence level α. The *subjective classical estimator* is $\boldsymbol{\theta}^*(\hat{\lambda}_T)$, where:

(1) if $\hat{z}_T(\boldsymbol{\theta}^*(0)) \le \eta_{\alpha,k}$, $\hat{\lambda}_T = 0$;
(2) if $\hat{z}_T(\boldsymbol{\theta}^*(0)) > \eta_{\alpha,k}$, $\hat{\lambda}_T$ is the solution to the following constrained maximization problem:

$$\max_{\lambda\in[0,1]} \hat{U}_T(\boldsymbol{\theta}^*(\lambda))$$
$$\text{s.t. } \hat{z}_T(\boldsymbol{\theta}^*(\lambda)) = \eta_{\alpha,k}. \tag{8}$$

This estimator generalizes the estimator found in Equation (5) proposed in Section 3 to any objective function satisfying Conditions 1–8 and to any dimension of the decision variable $\boldsymbol{\theta}$. It first checks whether the given default decision $\tilde{\boldsymbol{\theta}}$ is supported by the data. If for given $\tilde{\boldsymbol{\theta}}$ and confidence level α, the objective function cannot be increased in a statistically significant way, the default decision $\tilde{\boldsymbol{\theta}}$ is retained as the forecast estimator. Rejection of the null hypothesis implies that it is possible to move away from $\tilde{\boldsymbol{\theta}}$ and increase (in a statistical sense) the objective function. The classical estimate $\hat{\boldsymbol{\theta}}_T$ (representing the maximum of the empirical equivalent of the objective function) provides the natural direction towards which to move. The new estimator is, therefore, obtained by shrinking the default decision $\tilde{\boldsymbol{\theta}}$ towards the classical estimate $\hat{\boldsymbol{\theta}}_T$. The amount of shrinkage is determined by the constraint in Equation (8) and is given by the point where the increase in the objective function stops being statistically significant. This happens at the boundary of the confidence set.

Notice that, as stated at the end of Remark 3 in Section 3, the constrained maximization problem found in Equation (8) is equivalent to $\min_{\lambda\in[0,1]} \lambda$, s.t. $\hat{z}_T(\boldsymbol{\theta}^*(\lambda)) \le \eta_{\alpha,k}$.

The following theorem shows that the new estimator is consistent and establishes its relationship with the classical estimator.

Theorem 1 (Properties of the subjective classical estimator). Under Conditions 1–8 the subjective classical estimator $\boldsymbol{\theta}^*(\hat{\lambda}_T)$ of Definition 2 satisfies the following properties:

1. If $\alpha = 1$, $\boldsymbol{\theta}^*(\hat{\lambda}_T)$ is the classical estimator.
2. If $\alpha > 0$, $\boldsymbol{\theta}^*(\hat{\lambda}_T)$ is consistent.

Proof. See Appendix.

The intuition behind this result is that as the sample size grows the distribution of the first derivatives will be more and more concentrated around its true mean. Since, according to Definition 2, the estimator cannot be outside the perimeter of the $1-\alpha$ confidence interval, Conditions 1–8 guarantee that the confidence interval shrinks asymptotically towards zero and, therefore, the consistency of the estimator.

5.1 Relationship With Bayesian Econometrics

An important issue that deserves discussion is the relationship between the subjective classical estimator and the Bayesian approach to incorporating judgment. Bayesian estimators require the specification of a prior probability distribution $\pi(\boldsymbol{\theta})$. This prior distribution is then updated with the information contained in the sample, by applying the Bayes' rule. The posterior density of $\boldsymbol{\theta}$ given the sample data $y^T \equiv \{y_t\}_{t=1}^T$ is given by:

$$\pi(\boldsymbol{\theta}|y^T) = \frac{f(y^T|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{m(y^T)},$$

where $f(y^T|\boldsymbol{\theta})$ denotes the sampling distribution and $m(y^T)$ the marginal distribution of y^T. In the Bayesian framework, nonsample information is incorporated in the econometric analysis via the prior $\pi(\boldsymbol{\theta})$. Once the prior has been formulated, Bayesian techniques can be applied to find the $\hat{\boldsymbol{\theta}}_T^B$ which maximizes the expected utility (see Berger 1985 and the references cited in Granger and Machina 2006 for further discussion).

The estimator proposed in this article offers a classical alternative to Bayesian techniques to account for nonsample information in forecasting. The nonsample information is summarized by the two parameters $(\tilde{\boldsymbol{\theta}}, \alpha)$. $\tilde{\boldsymbol{\theta}}$ is the agent's default decision, while α is the confidence level in such a decision, which is then used to test the hypothesis $H_0: \tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}^0$, where $\boldsymbol{\theta}^0$ represents the optimal decision variable.

There is one special case in which the two estimators coincide. This happens when the decision maker is absolutely certain about the default decision $\tilde{\boldsymbol{\theta}}$. In this case, the prior distribution collapses to a degenerate distribution with total mass on the point $\tilde{\boldsymbol{\theta}}$. With such a prior, the posterior will always be identical to the prior, no matter what the sample data looks like and the Bayesian estimator will be $\hat{\boldsymbol{\theta}}_T^B = \tilde{\boldsymbol{\theta}}$. In our setting, on the other hand, certainty about parameter values may be expressed by setting $\alpha = 0$. When $\alpha = 0$ the null hypothesis can never be rejected and the new estimator becomes $\boldsymbol{\theta}^*(\hat{\lambda}_T) = \tilde{\boldsymbol{\theta}}$.

A second special case is when the decision maker has no information to exploit, besides that incorporated in the sample. Although this is a highly controversial issue in Bayesian statistics (see for instance Poirier 1995, pp. 321–331, and the references therein), lack of nonsample information can be accommodated in the Bayesian framework by choosing a diffuse prior. In some special cases—for example when estimating the mean of a Gaussian distribution with known variance—the Bayesian estimator based on a normal prior with variance tending to infinity is known to converge to the classical estimator. In our setting, instead, lack of information can be easily incorporated by setting $\alpha = 1$, which, as shown in Theorem 1, leads to the classical estimator.

For the intermediate cases, there is no obvious mapping between Bayesian priors and our subjective parameters $(\tilde{\boldsymbol{\theta}}, \alpha)$. The choice between the two estimators depends on how the nonsample information is formalized. Our estimator is not suited to exploit nonsample information which takes the form of a prior probability distribution, and in this case one needs to resort to Bayesian estimation procedures. On the other hand, application of Bayes' rule relies on the availability of a fully specified prior probability distribution and if the nonsample information is expressed in terms of the two parameters $(\tilde{\boldsymbol{\theta}}, \alpha)$, one needs to use the subjective classical estimator. From this perspective, the choice between Bayesian and classical econometrics is not an issue to be settled by econometricians, but rather by the decision maker through the format in which she/he provides the nonsample information.

A major potential problem with Bayesian econometrics is that formulation of priors is not always obvious and may, in some cases, put a heavy burden on the decision maker. As an example of such difficulties, consider again the risk analysis of Section 4. In Figure 2, we compare the risk function of the subjective classical estimator with that of two Bayesian estimators, one based on a normal prior and one based on a Laplace (also known as double exponential) prior. We calibrated the priors in such a way that all risk functions intersect at the same point. We see that these two specific Bayesian estimators are dominated by the subjective classical estimator for low and high values of $\boldsymbol{\theta}^0$, while the converse is true over a finite intermediate interval. The risk function of the Laplace estimator has an upper bound at around 3.8, while the subjective classical estimator has an upper bound at around 3.7.

In Figure 3, we report the calibrated priors. The Laplace distribution has greater mass around zero and thicker tails with respect to the normal distribution. The tail behavior of the prior distribution leads to estimators with substantially different properties, as shown in Figure 2. An undesirable characteristic of the normal Bayesian estimator is that its risk function diverges to infinity for increasing values of $\boldsymbol{\theta}^0$: there is no limit to the potential damage inflicted by poorly chosen priors. As discussed in Magnus (2002), the unbounded risk is a consequence of the fact that the tails of the normal distribution are too thin. The heavy tails of the Laplace distribution on the other hand guarantee a bounded risk function. The shape of the risk functions of the subjective classical estimator on the other hand is not particularly sensitive to small changes of α, as illustrated in Figure 4.

How difficult it is to elicit the tail behavior of the priors depends on the knowledge of the decision maker and on the type of problems she/he is facing. These difficulties would be further exacerbated in a multivariate context. The examples discussed in the next section show that the subjective classical estimator instead is relatively straightforward to implement.

6. EXAMPLES

We illustrate with two examples how the theory developed in the previous section can be implemented.

In the first example, we estimate the optimal portfolio weights maximizing a mean-variance utility function. We highlight how the theory proposed in this article naturally takes into account the impact of estimation errors and show, with an out of sample exercise, how the subjective classical estimator outperforms a given benchmark portfolio.

The second example is an application to U.S. GDP forecast. We show how one can map a default decision on future GDP growth rates into default parameters of the econometrician's favorite model. We provide an illustration using an autoregressive model to forecast quarterly GDP growth rates.

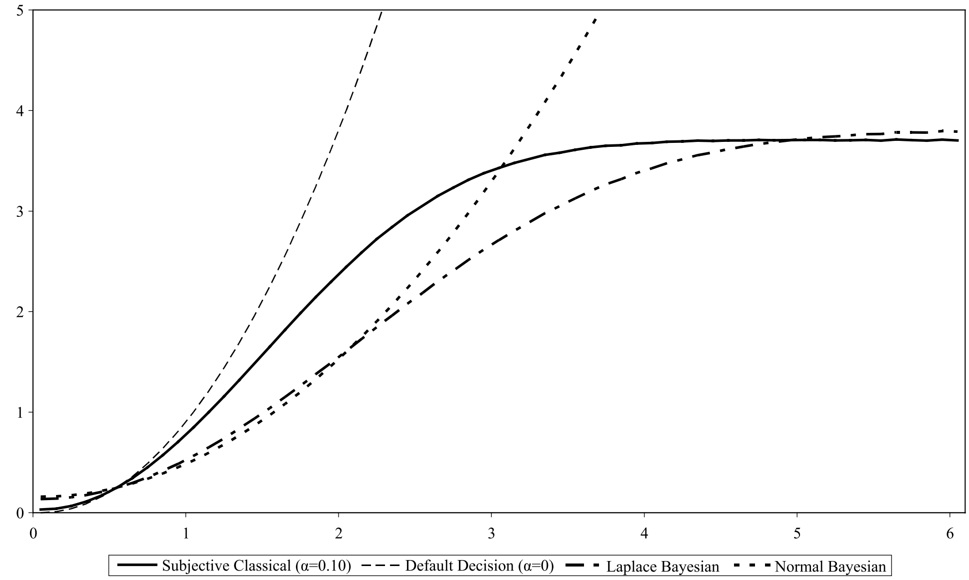

Figure 2. Risk functions of Bayesian estimators with different priors.

6.1 Mean-Variance Asset Allocation

Markowitz's (1952) mean-variance model provides the standard benchmark for portfolio allocation. It formalizes the intuition that investors optimize the trade off between returns and risks, resulting in optimal portfolio allocations which are a function of expected return, variance (the proxy used for risk), and the degree of risk aversion of the decision maker. Despite its theoretical appeal, it is well known that standard implementations of this model produce portfolio allocations with no economic intuition and little (if not negative) investment value. These problems were initially pointed out, among others, by Jobson and Korkie (1981), who used a Monte Carlo experiment to show that estimated mean-variance frontiers can be quite far away from the true ones. The crux of the problem is colorfully, but effectively, highlighted by the following quotation of Michaud (1998, p. 3):

> "[Mean-variance optimizers] overuse statistically estimated information and magnify the impact of estimation errors. It is not simply a matter of garbage in, garbage out, but, rather, a molehill of garbage in, a mountain of garbage out."

The problem can be restated in terms of the theory developed in Section 5. Classical estimators maximize the empirical expected utility, without taking into consideration whether they are statistically significantly better than the default decision. Our theory provides a natural alternative. For a given benchmark portfolio (the default decision $\tilde{\boldsymbol{\theta}}$ in the notation of Section 5) and a confidence level α, the resulting optimal portfolio is the one which increases the empirical expected utility as long as the first derivatives are statistically different from zero.

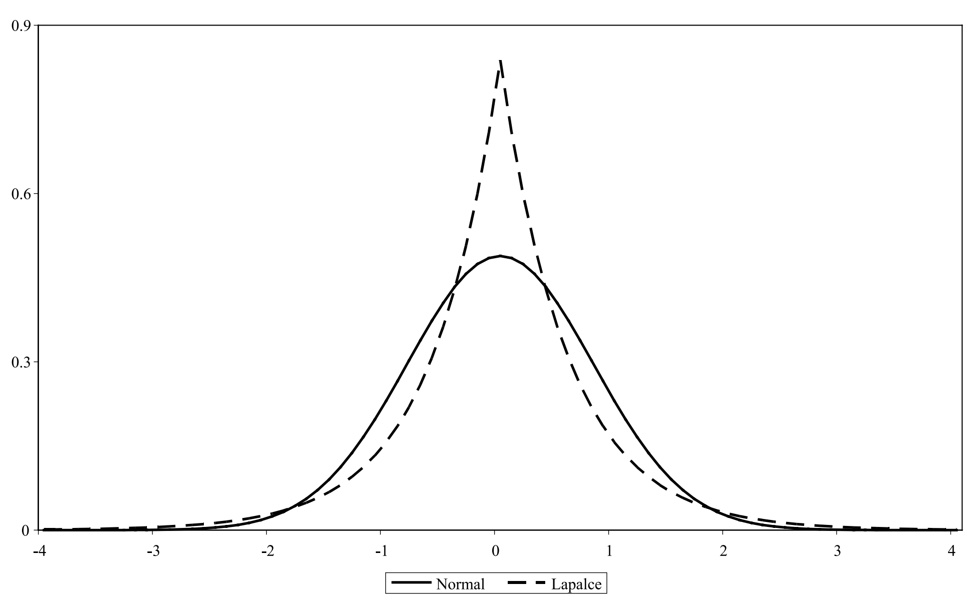

Figure 3. Normal and Laplace prior distributions.

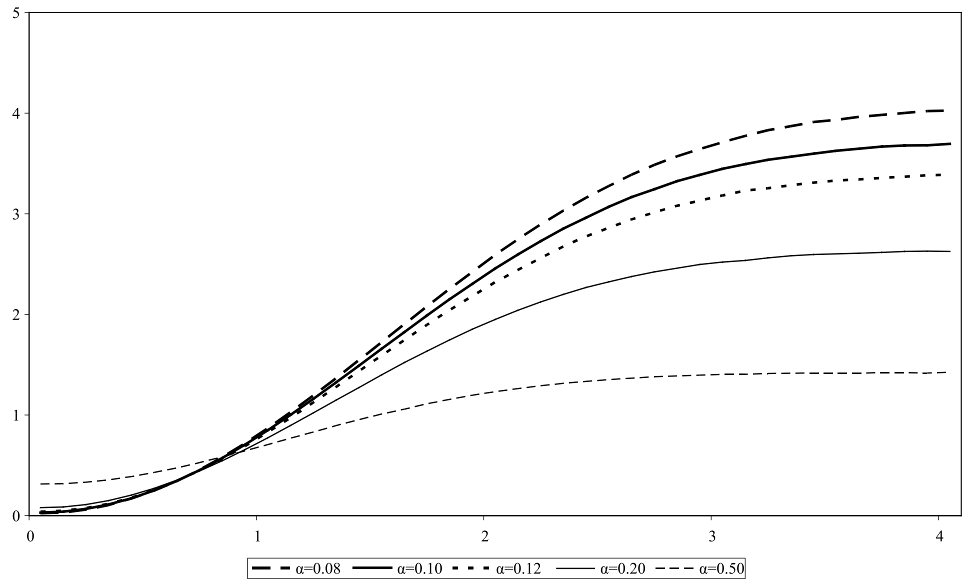

Figure 4. Sensitivity of subjective classical estimators to different αs.

To formalize this discussion, consider a portfolio with $N+1$ assets. Denote with $\boldsymbol{\theta}$ the N-vector of weights associated to the first N assets entering a given portfolio, and denote with $y_t(\boldsymbol{\theta})$ the portfolio return at time t, where the dependence on the individual asset weights has been made explicit. Since all the weights must sum to one, note that $\theta_{N+1} = 1 - \sum_{i=1}^{N} \theta_i$, where θ_i denotes the ith element of $\boldsymbol{\theta}$ and θ_{N+1} is the weight associated to the $(N+1)$th asset of the portfolio. Let's assume an investor wants to maximize a trade-off between mean and variance of portfolio returns, resulting in the following expected utility function:

$$\begin{aligned} U(\boldsymbol{\theta}) &= E[y_{T+1}(\boldsymbol{\theta})] - \xi V[y_{T+1}(\boldsymbol{\theta})] \\ &= E[y_{T+1}(\boldsymbol{\theta})] - \xi \{E[y_{T+1}^2(\boldsymbol{\theta})] - E[y_{T+1}(\boldsymbol{\theta})]^2\}, \qquad (9) \end{aligned}$$

where ξ describes the investor's attitude towards risk. The empirical analogue is:

$$\begin{aligned} \hat{U}_T(\boldsymbol{\theta}) &= T^{-1} \sum_{t=1}^{T} y_t(\boldsymbol{\theta}) \\ &\quad - \xi \left\{ T^{-1} \sum_{t=1}^{T} y_t^2(\boldsymbol{\theta}) - \left[T^{-1} \sum_{t=1}^{T} y_t(\boldsymbol{\theta}) \right]^2 \right\}. \qquad (10) \end{aligned}$$

The first order conditions are:

$$\begin{aligned} \nabla_\theta \hat{U}(\boldsymbol{\theta}) &= T^{-1} \sum_{t=1}^{T} \nabla_\theta y_t(\boldsymbol{\theta}) \\ &\quad - \xi \left\{ T^{-1} 2 \sum_{t=1}^{T} y_t(\boldsymbol{\theta}) \nabla_\theta y_t(\boldsymbol{\theta}) \right. \\ &\quad \left. - 2 \left[T^{-1} \sum_{t=1}^{T} y_t(\boldsymbol{\theta}) \right] T^{-1} \sum_{t=1}^{T} \nabla_\theta y_t(\boldsymbol{\theta}) \right\}, \end{aligned}$$

where $\nabla_\theta y_t(\boldsymbol{\theta}) \equiv \mathbf{y}_t^N - y_t^{N+1} \boldsymbol{\iota}$, $\mathbf{y}_t^N$ is an N-vector containing the returns at time t of the first N assets, y_t^{N+1} is the return at time t of the $(N+1)$th asset, and $\boldsymbol{\iota}$ is an N-vector of ones. The variance-covariance matrix of the first derivatives is computed using the outer product estimate.

We apply the methodology developed in Section 5 to monthly log returns of the stocks composing the Dow Jones Industrial Average (DJIA) index, as of July 15, 2005. The sample runs from January 1, 1987 to July 1, 2005, for a total of 225 observations. We set $\xi = 1$ and use as default decision the equally weighted portfolio and confidence levels $\alpha = 1, 0.10, 0.01$. Although any other benchmark could be used, the equally weighted portfolio has emerged as a natural benchmark in the literature. For instance, DeMiguel, Garlappi, and Uppal (2009) showed that, among 14 estimated models, none of them is consistently better than the naive $1/N$ portfolio. Regarding the choice of the confidence level α, recall from Figures 1 and 4 that it reflects the risk of underperforming with respect to the benchmark portfolio: the lower α, the lower such risk. At the same time the lower α, the lower the improvement over the benchmark, when the benchmark is far away from the optimal portfolio. This dichotomy corresponds to the probability of Type I error (falsely rejecting the null) and Type II error (falsely accepting the null) in hypothesis testing. For a fixed sample, it is not possible to simultaneously reduce both probabilities and the decision maker faces an inevitable trade-off. In the following example we will use $\alpha = 0.01$ and 0.10, which are values typically used in hypothesis testing.

Notice that the case with $\alpha = 1$ corresponds to the standard implementation of the mean-variance model, i.e., it corresponds to the case where the sample estimates of expected returns and variance-covariances are substituted into the analytical solution of the optimal portfolio weights.

We recursively estimated the optimal weights associated to the different confidence levels α for portfolios with a different number of assets, namely, 4, 16, and 30. Following DeMiguel, Garlappi, and Uppal (2009), we evaluate the out of sample performance of the estimators using rolling windows of $M = 60$ and $M = 120$ observations. That is, at each month t, starting

Table 1. Average difference in out of sample realized utilities associated to $\boldsymbol{\theta}^*(\hat{\lambda}_T)$ and to the equal weighted benchmark. p-values of the Giacomini and White (2006) test of predictive ability in parenthesis. Values significant at the 10%, 5%, and 1% levels are denoted by one, two, and three asterisks, respectively

	$M=60$			$M=120$		
	$N=4$	$N=16$	$N=30$	$N=4$	$N=16$	$N=30$
$\alpha=1$	−0.30	1.00	-8.40^{***}	1.82	6.47^{*}	2.61
	(0.82)	(0.71)	(0.01)	(0.37)	(0.08)	(0.45)
$\alpha=0.10$	0.48^{*}	0.09	0	0.82^{*}	4.19^{***}	1.48^{*}
	(0.06)	(0.59)	–	(0.07)	(0.01)	(0.06)
$\alpha=0.01$	0.02	0	0	0.01	0.92^{**}	0
	(0.49)	–	–	(0.71)	(0.03)	–

from $t=M$, we estimate the optimal weights using the previous M observations. Next, we compute the out of sample realized return at $t+1$ of the optimal portfolio. With a sample of size T, we obtain a total of $T-M$ out of sample observations. Finally, we compute average realized utilities associated to the optimal portfolio out of sample returns. In Table 1 we report the difference in average utilities between the optimal portfolio and the equal weight benchmark. That is, if we denote with $U_t(\boldsymbol{\theta}) \equiv y_t(\boldsymbol{\theta}) - \xi\{y_t^2(\boldsymbol{\theta}) - [(T-M)^{-1}\sum_{t=M+1}^{T} y_t(\boldsymbol{\theta})]y_t(\boldsymbol{\theta})\}$ the realized utility associated to portfolio $\boldsymbol{\theta}$, we compute:

$$Z_{T-M} \equiv (T-M)^{-1}\sum_{t=M+1}^{T}\left[U_t(\boldsymbol{\theta}^*(\hat{\lambda}_{t-1})) - U_t(\tilde{\boldsymbol{\theta}})\right]. \qquad (11)$$

We also compute the test of predictive ability associated to the statistic Z_{T-M}, as suggested by Giacomini and White (2006). We report in parenthesis the p-values associated to the null hypothesis that the two portfolios have equal performance.

Let's look first at the performance of the estimators with a window of 60 observations. We notice that the classical estimator ($\alpha=1$) performs worse than the equal weight benchmark for portfolios with 4 and 30 assets. In the case of the 30 asset portfolio, the inferior performance of the classical estimator is also statistically significant at the 1% confidence level. As we decrease α, the performance of the estimator improves. With $N=4$, we see that the optimized portfolio outperforms the benchmark in a statistically significant way for $\alpha=0.10$. When $\alpha=0.01$, the difference in realized expected utilities is still positive but no longer statistically significant. For the portfolio with 16 assets, none of the optimized portfolios significantly outperforms the benchmark. For $N=30$, the benchmark is not rejected at the 10% level. (Note that the zeros in the table reflect the fact that the benchmark portfolio is never rejected. In this case $\boldsymbol{\theta}_T^* = \tilde{\boldsymbol{\theta}}$ and, therefore, the optimized and benchmark portfolios are identical.)

With an estimation window of 120 observations, the performance of all estimators improves. With a larger estimation sample, estimates become more precise. It is nevertheless interesting to notice that the performance of the classical estimator can be improved upon, by reducing the size of α. With $N=4$, the outperformance of the classical estimator is not statistically significant, while it is with $\alpha=0.10$. With $N=16$, the outperformance of the estimator with $\alpha=1$ is statistically significant at the 10% level, but one can increase its significance by reducing α to 0.10. Finally, with $N=30$, the difference in realized expected utilities is significant only by choosing $\alpha=0.10$.

What emerges from these results is that the smaller M and the larger N, the greater the impact of estimation error on the optimal portfolio weights. Consistently with the results discussed in Section 4, the lower α, the more confident the decision maker can be that the resulting allocation will beat the benchmark portfolio. There is, however, no free lunch: choosing too small an α will result in more conservative portfolios (i.e., portfolios closer to the benchmark), implying that the decision maker may forgo potential increases in the expected utility. This can be linked back to the discussion of Figure 1 in Section 4. The lower α, the higher the likelihood that the subjective classical estimator will have lower risk than the default decision (i.e., the risk function of the subjective classical estimator will lie below that of the default decision, except for values of $\boldsymbol{\theta}^0$ very close to $\tilde{\boldsymbol{\theta}}$). At the same time, the lower α, the higher the risk associated to the subjective classical estimator for values of $\boldsymbol{\theta}^0$ far away from $\tilde{\boldsymbol{\theta}}$.

6.2 Forecasting U.S. GDP

A possible difficulty in implementing the estimator of Section 5 is related to the formulation of a default decision in terms of the parameters of an econometric model about which the decision maker may know nothing or very little. We propose a simple strategy to map a default decision on the variable of interest to the decision maker (GDP in this case) into default decisions on the parameters of the econometrician's favorite model.

In principle, it is possible to express a default decision directly on the parameter vector $\boldsymbol{\theta}$ or indirectly on the dependent variable y_{T+1} to be forecast. If the decision maker can formulate a guess on $\boldsymbol{\theta}$, the theory of Section 5 can be applied directly. In most circumstances, however, it may be more natural to have a judgment about the future behavior of y_{T+1}, rather than about abstract model parameters. Let's denote this default decision as $\tilde{y}_{T+1}$. Using the notation of Section 5, this can be translated into default parameters on $\boldsymbol{\theta}$ as follows:

$$\tilde{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \hat{U}_T(\boldsymbol{\theta})$$
$$\text{s.t. } \hat{y}_{T+1}(\boldsymbol{\theta}) = \tilde{y}_{T+1}, \qquad (12)$$

where $\hat{y}_{T+1}(\boldsymbol{\theta})$ is the model's forecast conditional on the parameter vector $\boldsymbol{\theta}$. The default decision $\tilde{y}_{T+1}$ is mapped into a default parameter vector by choosing the $\tilde{\boldsymbol{\theta}}$ that maximizes the objective function subject to the constraint that the forecast at time T is equal to $\tilde{y}_{T+1}$.

Let's consider, for concreteness, an application to quarterly GDP forecasting, using an AR(4) model:

$$y_t = \theta_0 + \sum_{i=1}^{4}\theta_i y_{t-i} + \varepsilon_t. \qquad (13)$$

If the decision maker has a quadratic loss function, we have $\hat{U}_T(\boldsymbol{\theta}) \equiv -T^{-1}\sum_{t=1}^{T}[y_t - \hat{y}_t(\boldsymbol{\theta})]^2$, where $\hat{y}_t(\boldsymbol{\theta}) \equiv \theta_0 + \sum_{i=1}^{4}\theta_i y_{t-i}$. The score evaluated at $\tilde{\boldsymbol{\theta}}$ is

$$\nabla_\theta \hat{U}_T(\tilde{\boldsymbol{\theta}}) = 2T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t(\tilde{\boldsymbol{\theta}})\nabla_\theta \hat{y}_t(\tilde{\boldsymbol{\theta}}), \qquad (14)$$

Table 2. Subjective guesses $\tilde{\boldsymbol{\theta}}$ and estimated parameters $\boldsymbol{\theta}^*_T(\hat{\lambda})$ associated to different subjective guesses on Q4 2005 GDP growth rates (3% and 5%), with $\alpha = 0.10$. A subjective guess of 3% is not rejected by the data and maps into parameter values very close to the OLS $\hat{\boldsymbol{\theta}}_T$. A subjective guess of 5%, instead, is rejected by the data, resulting in parameter estimates different from the parameter guess

		$\tilde{y}_{T+1} = 3\%$		$\tilde{y}_{T+1} = 5\%$	
	$\hat{\boldsymbol{\theta}}_T$	$\tilde{\boldsymbol{\theta}}$	$\boldsymbol{\theta}^*(\hat{\lambda}_T)$	$\tilde{\boldsymbol{\theta}}$	$\boldsymbol{\theta}^*(\hat{\lambda}_T)$
θ_0	1.65	1.54	1.54	2.73	1.99
θ_1	0.23	0.21	0.21	0.42	0.29
θ_2	0.36	0.37	0.37	0.30	0.34
θ_3	−0.16	−0.18	−0.18	−0.03	−0.12
θ_4	0.04	0.05	0.05	−0.04	0.01
$\hat{y}(\theta)$	3.19%	3%	3%	5%	3.77%

where $\hat{\varepsilon}_t(\tilde{\boldsymbol{\theta}}) \equiv y_t - \hat{y}_t(\tilde{\boldsymbol{\theta}})$ and $\nabla_\theta \hat{y}_t(\tilde{\boldsymbol{\theta}}) \equiv [1, y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}]'$. We estimate the asymptotic variance-covariance matrix of the score using standard heteroscedasticity-consistent estimators (White 1980):

$$\hat{\boldsymbol{\Sigma}}_T \equiv 4T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t(\hat{\boldsymbol{\theta}}_T)^2 \nabla_\theta \hat{y}_t(\hat{\boldsymbol{\theta}}_T) \nabla'_\theta \hat{y}_t(\hat{\boldsymbol{\theta}}_T), \qquad (15)$$

where $\hat{\boldsymbol{\theta}}_T$ is the OLS estimate. We estimate this model using quarterly data for the U.S. real GDP growth rates. The data are taken from the FRED® database (see *http://research.stlouisfed.org/fred2*). The data has been seasonally adjusted and our sample runs from Q1 1983 to Q3 2005, with 90 observations. The growth rates are computed as log differences.

For illustratory purposes, we consider two different default decisions for GDP growth in the next quarter (Q4 2005), $\tilde{y}_{T+1} = 3\%$ and $\tilde{y}_{T+1} = 5\%$, both with a confidence level $\alpha = 0.10$. The results reported in Table 2 show that $\tilde{y}_{T+1} = 3\%$ maps into a parameter guess $\tilde{\boldsymbol{\theta}}$ which cannot be rejected by the data $[\boldsymbol{\theta}^*(\hat{\lambda}_T) = \tilde{\boldsymbol{\theta}}]$. These parameter values are also very close to the OLS estimates $\hat{\boldsymbol{\theta}}_T$, resulting in very similar forecasts. Note that in this case the forecast associated to $\boldsymbol{\theta}^*(\hat{\lambda}_T)$ is equal to 3%, the original default decision ($\tilde{y}_{T+1} = 3\%$).

The other default decision, $\tilde{y}_{T+1} = 5\%$, is instead rejected by the data at the chosen confidence level, resulting in parameter estimates $\boldsymbol{\theta}^*(\hat{\lambda}_T)$ which are different from the parameter guess $\tilde{\boldsymbol{\theta}}$. The estimated shrinkage factor $\hat{\lambda}_T$ was 0.68. The out of sample GDP forecast at Q4 2005 associated to $\boldsymbol{\theta}^*(\hat{\lambda}_T)$ is 3.77% and the OLS forecast is 3.19%, both definitely lower than the default decision of 5%.

7. CONCLUSION

Classical forecasts typically ignore nonsample information and estimation errors due to finite sample approximations. In this article we pointed out how these two problems are connected. We explicitly introduced into the classical estimation framework two elements: a default decision and a confidence associated to it. Their role is to explicitly take into consideration the nonsample information available to the decision maker. These elements served to define a new estimator, which increases the objective function only as long as the improvement is statistically significant, and to formalize the interaction between judgment and data in the forecasting process.

We illustrated with a detailed risk analysis the properties of the new estimator. We provided two applications, which give strong support to our theory. We illustrated how our new estimator may provide a satisfactory solution to the well-known implementation problems of the mean-variance asset allocation model. We also showed how a default decision on the variable to be forecast can be mapped into default parameters of the econometrician's favorite model.

APPENDIX

Proof of Theorem 1 (Properties of the New Estimator)

1. If $\alpha = 1$, $\eta_{\alpha,k} = 0$ and the constraint in Equation (8) becomes $\hat{z}_T(\boldsymbol{\theta}^*(\lambda)) = 0$. This implies $\nabla_\theta \hat{U}_T(\boldsymbol{\theta}^*(\hat{\lambda}_T)) = 0$, which coupled with Equation (8) implies $\boldsymbol{\theta}^*(\hat{\lambda}_T) = \hat{\boldsymbol{\theta}}_T$, where $\hat{\boldsymbol{\theta}}_T$ is defined in Definition 1.
2. Let $\boldsymbol{\theta}^0 \equiv \text{plim}_{T\to\infty} \hat{\boldsymbol{\theta}}_T$. We need to show that $\boldsymbol{\theta}^*(\hat{\lambda}_T) \xrightarrow{p} \boldsymbol{\theta}^0$. By Equation (8) and Conditions 2 and 8, this is equivalent to showing that $\|\nabla_\theta \hat{U}_T(\boldsymbol{\theta}^*(\hat{\lambda}_T))\| \xrightarrow{p} 0$. Note that Definition 2 implies $\Pr(\hat{z}_T(\boldsymbol{\theta}^*(\hat{\lambda}_T)) > \eta_{\alpha,k}) = 0$. Conditions 1, 2, 3, and 6 imply that $\|\nabla_\theta \hat{U}_T(\boldsymbol{\theta}^*(\hat{\lambda}_T))\| \xrightarrow{p} c$. Suppose by contradiction that $c \neq 0$. Then $\Pr(\hat{z}_T(\boldsymbol{\theta}^*(\hat{\lambda}_T)) > \eta_{\alpha,k}) = \Pr(\nabla'_\theta \hat{U}_T(\boldsymbol{\theta}^*(\hat{\lambda}_T))\hat{\boldsymbol{\Sigma}}_T^{-1} \nabla_\theta \hat{U}_T(\boldsymbol{\theta}^*(\hat{\lambda}_T)) > \eta_{\alpha,k}/T)$. But since $\nabla'_\theta \hat{U}_T(\boldsymbol{\theta}^*(\hat{\lambda}_T))\hat{\boldsymbol{\Sigma}}_T^{-1} \nabla_\theta \hat{U}_T(\boldsymbol{\theta}^*(\hat{\lambda}_T))$ is bounded in probability above zero and $\eta_{\alpha,k}/T$ converges to 0 as T goes to infinity, for any $q \in [0, 1)$ there must exist a T^* such that, for any $T > T^*$, $\Pr(\nabla'_\theta \hat{U}_T(\boldsymbol{\theta}^*(\hat{\lambda}_T))\hat{\boldsymbol{\Sigma}}_T^{-1} \times \nabla_\theta \hat{U}_T(\boldsymbol{\theta}^*(\hat{\lambda}_T)) > \eta_{\alpha,k}/T) > q$. This implies a violation of the constraint in Equation (8) and, therefore, a contradiction.

ACKNOWLEDGMENTS

I would like to thank, without implicating them, two anonymous referees, Lorenzo Cappiello, Matteo Ciccarelli, Rob Engle, Lutz Kilian, Jan Magnus, Benoit Mojon, Cyril Monnet, Andrew Patton and Timo Teräsvirta for their comments and useful suggestions. I also would like to thank the ECB Working Paper Series editorial board and seminar participants at the ECB, Tilburg, Federal Reserve Board of Governors, Federal Reserve Bank of Chicago, Oxford, Stanford and UCSD. The views expressed in this paper are those of the author and do not necessarily reflect those of the European Central Bank or the Eurosystem.

[Received February 2008. Revised May 2009.]

REFERENCES

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

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

Durlauf, S. N. (2002), Discussion of "The Role of Models and Probabilities in the Monetary Policy Process," by C. Sims, *Brooking Papers on Economic Activity*, 2, 41–50.

Giacomini, R., and White, H. (2006), "Test of Conditional Predictive Ability," *Econometrica*, 74, 1545–1578.

Gonzalez, A., Hubrich, K., and Teräsvirta, T. (2009), "Forecasting Inflation With Gradual Regime Shifts and Exogenous Information," Research Paper 2009-3, CREATES.

Granger, C. W. J., and Machina, M. (2006), "Forecasting and Decision Theory," in *Handbook of Economic Forecasting*, eds. G. Elliott, C. W. J. Granger, and A. Timmermann, Amsterdam: North-Holland, pp. 81–98.

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

Haavelmo, T. (1944), "Supplement to 'The Probability Approach in Econometrics'," *Econometrica Supplemental Material*, 12.

Hansen, P. R. (2007), "In-Sample and Out-of-Sample Fit: Their Joint Distribution and Its Implications for Model Selection and Model Averaging (Criteria-Based Shrinkage for Forecasting)," mimeo, Stanford University.

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

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

Magnus, J. R., and Durbin, J. (1999), "Estimation of Regression Coefficients of Interest When Other Regression Coefficients Are of No Interest," *Econometrica*, 67, 639–643.

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.

Newey, W., and McFadden, D. (1994), "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, Vol. IV, eds. R. F. Engle and D. L. McFadden, Amsterdam: Elsevier Science.

Poirier, D. J. (1995), *Intermediate Statistics and Econometrics. A Comparative Approach*, Cambridge, MA: MIT Press.

White, H. (1980), "A Heteroskedaticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.