Forecasting with Judgment

Simone Manganelli*

15 October 2007

Abstract

This paper argues that forecasts should maximise the objective function in a stochastic, rather than deterministic, way. We explicitly incorporate two elements into the estimation framework: a subjective guess on the variable to be forecast and a probability reflecting the confidence associated with it. Starting from the subjective guess, the judgmental forecast increases the objective function only as long as its first derivatives are statistically different from zero. We show that the new estimator includes as a special case the classical estimator and discuss its relationship with Bayesian estimators. We study the properties of this new estimator with a detailed risk analysis and a Monte Carlo simulation. Finally, we illustrate the performance of the estimator with applications to mean-variance portfolio selection and to GDP forecast.

*I would like to thank, without implicating them, Lorenzo Cappiello, Matteo Ciccarelli, Rob Engle, Lutz Kilian, Benoit Mojon, Cyril Monnet, Andrew Patton and Timo Teräsvirta for their encouragement and useful suggestions. I also would like to thank the ECB Working Paper Series editorial board and seminar participants at the ECB, the 2006 European Meeting of the Econometric Society (Vienna), the 2006 meeting of the European Finance Association (Zurich), and the 7^{th} macroeconometric workshop at the Halle Institute for Economic Research. DG-Research, European Central Bank, Kaiserstrasse 29, D-60311 Frankfurt am Main, Germany. E-mail: simone.manganelli@ecb.int.

Keywords: Decision under uncertainty, estimation, overfitting, asset allocation.

JEL classification: C13, C53, G11.

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. The classical theory of forecasting builds on the assumption that agents wish to minimise the expected cost associated with these errors (see chapter VI of Haavelmo 1944, Granger and Newbold 1986, Granger and Machina 2005). Classical forecasts are obtained as the minimisers of the sample equivalent of the unobservable expected cost.

We emphasise two closely related problems with this theory. First, classical forecasts do not explicitly account for non-sample information available to the decision-maker, even though subjective judgment often plays an important role in many real world forecasting processes. Second, classical estimators maximise an objective function, which depends on unknown parameters and is only a finite sample approximation of the true function to be maximised. Since any finite sample approximation is subject to estimation error, maximisation of this function does not necessarily coincide with the maximisation of the true objective function.

To address the first problem, we explicitly introduce two elements in the estimation framework. The first element is a subjective guess on the variable to be forecast. The second is a probability reflecting the confidence of the decision-maker in this guess. These elements serve to summarise the non-sample information available to the decision-maker, and allow us to formalise the interaction between judgment and data in the forecasting process.

As for the second problem, estimation error is explicitly taken into consideration by testing whether the subjective guess statistically maximises the objective function. This is equivalent to testing whether its first derivatives evaluated at the parameter values implied by the subjective guess are statistically equal to zero at the given confidence level. If this is the case, the subjective guess is retained as the estimate. Otherwise, the objective function is increased as long as its first derivatives are significantly different from zero, and a new, model-based forecast is obtained. The resulting estimator will be on the perimeter of the confidence set, unless the subjective guess is within the confidence set, in which case the estimator equals the subjective guess.

Classical estimators deterministically 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 any subjective guess. 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 subjective guess becomes less and less relevant.

Subjective guess 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 paper. The classical procedure first derives the statistical properties of the estimator and then tests whether the estimate is different from the subjective guess, for the pre-specifed confidence level. The problem with this procedure is that it is unclear what to do when the null is rejected. In the procedure proposed in this paper, the null hypothesis is that the first derivatives of the objective function evaluated at the subjective guess are equal to zero, while the alternative hypothesis is that they are not. If the null is rejected, the alternative tells that at the margin 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 subjective guess is 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 subjective guess are statistically different from zero, and therefore that the objective function can be statistically improved. However, the objective function can be improved *only* up to the point where the first derivatives are no longer statistically different from zero. Beyond that point, and in particular at the point represented by the classical estimator, the probability of deteriorating, rather than improving, the objective function becomes greater than the chosen confidence level.

We illustrate this point with a detailed risk analysis of the proposed estimator, comparing its performance to the classical and pretest estimators. We show how our estimator, differently from the classical and pretest estimators, outperforms any given subjective guess in a precise statistical sense.

We also perform a Monte Carlo simulation with different sample sizes. We show that the new estimator outperforms the classical estimator, whenever the subjective guess is close enough to the true parameter value. We argue that the dichotomy between judgment and estimation implies that the forecasting process should be characterised by a clear separation between decision-makers — who should provide the judgment and their confidence in it — and econometricians, who should test whether such judgment is supported by the available data and models used for the analysis. It also implies that the responsibility for good or bad forecasts is shared between decision-makers and econometricians: having good judgment may be as important as the quality of the econometric model.

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 non-sample 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 the true model parameters) the two estimators coincide. In intermediate cases, the choice depends on how the non-sample information is formalised. Bayesian econometrics should be used whenever the non-sample information takes the form of a prior probability distribution. In contrast, the estimator proposed in this paper can be used whenever the non-sample information is expressed in terms of a subjective guess 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 s/he provides the non-sample 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 paper can be used to tackle some of the well-known implementation problems of mean-variance portfolio selection models. For a given benchmark portfolio (the subjective guess, 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 classical mean-variance optimisers.

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 subjective guess directly on these parameters. We suggest a simple and intuitive strategy to map the subjective guess on the variable of interest to the decision-maker into values for the parameters of the econometrician's favourite model. Specifically, these "judgmental parameter values" are obtained by maximising the objective function subject to the constraint that the forecast implied by the model is equal to the subjective guess. We illustrate how this works in the context of a simple autoregressive model.

The paper is structured as follows. In the next section, we use a stylised statistical model to highlight the problems associated to classical estimators. In section 3, we build on this stylised model to develop the heuristics behind the new forecasting theory. Section 4 presents a 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

In this section we illustrate with a simple example the problem associated with classical estimators. The intuition is the following. 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 i.i.d. 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 maximises the expected cost:

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

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 maximiser 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 don't maximise the expected utility, but also an error term $\varepsilon_T(\theta)$ which vanishes only asymptotically.

3 An Alternative Forecasting Strategy

The question is now whether we can find an alternative estimator which may have better properties than the classical one. To answer this question we introduce extra elements into the analysis: a subjective guess on the model's parameter and a probability summarising the confidence of the forecaster in this guess.

The first order condition of the optimal forecast problem (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 $\tilde{\theta}$ is some subjective guess and $\hat{y}_T \equiv T^{-1} \sum_{t=1}^{T} y_t$. $f_T(\tilde{\theta})$ 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, $f_T(\tilde{\theta}) \sim N(0, 1/T)$.

For a given confidence level α, let $\pm\kappa_{\alpha/2}$ denote the corresponding standard normal critical values and $\pm\eta_{\alpha/2}(T) \equiv \pm\sqrt{T^{-1}}\kappa_{\alpha/2}$. We suggest the following estimator:

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

That is, given the subjective guess $\tilde{\theta}$ and the confidence level α, if the null hypothesis $H_0 : f_T(\tilde{\theta}) = 0$ cannot be rejected, the subjective guess $\tilde{\theta}$ becomes the forecast. If the null is rejected, the forecast is given by the point where the first derivative $f_T(\theta_T^*)$ is exactly equal to its critical value. Note that $\eta_{\alpha/2}(T)$ converges to zero as T goes to infinity. Therefore θ_T^* converges to the classical estimator and is consistent.

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 subjective guess $\tilde{\theta}$ 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. S/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 (Non-sample information) - Both $\tilde{\theta}$ and α are exogenous to the statistical problem. They summarise the non-sample information available to the decision maker, and represent subjective elements in the analysis. The confidence level α may have different interpretations. It may be interpreted as the confidence of the forecaster in the subjective guess and in this case it should reflect the knowledge of the environment in which the forecast takes place. Alternatively, it may be thought of as the willingness to commit type I errors, i.e. of rejecting the null when $\tilde{\theta}$ is indeed the optimal forecast. Finally, since it determines when the increase in expected utility stops to be statistically significant, it may be seen as a device against overfitting. The forecaster will choose a low α whenever s/he is reasonably confident in the subjective guess $\tilde{\theta}$, and/or if the cost of committing type I errors is high, and/or if she is concerned about overfitting. Note that in the classical paradigm there is no place for subjective guesses and therefore $\alpha = 1$: in this case $\kappa_{\alpha/2} = 0$ and θ_T^* is simply the solution obtained by setting the first derivative (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 statistically different from zero, and therefore that the objective function can be increased in a statistically significant way. For the chosen confidence level α, the forecaster can statistically decrease the objective function *only* up to the point where the first derivative is no longer statistically different from zero, i.e. $f_T(\theta_T^*) = \eta_{\alpha/2}(T)$, if $f_T(\tilde{\theta}) > \eta_{\alpha/2}(T)$, or $f_T(\theta_T^*) = -\eta_{\alpha/2}(T)$, if $f_T(\tilde{\theta}) < -\eta_{\alpha/2}(T)$. Notice that for any point between θ_T^* and $\hat{\theta}_T$ the probability of decreasing (instead of increasing) the objective function in population is greater than α. It is only between $\tilde{\theta}$ and θ_T^* that the forecaster can be confident (in a statistical sense) that the objective function is actually increased.

Remark 4 (Relationship with the Burr estimator) - In the special example considered in this section, estimator (5) coincides with the Burr estimator, as discussed by Magnus (2002). Magnus arrived at this estimator following a completely different logic. He shows that the Burr estimator is the minimax regret estimator of a large class of estimators, defined by the class of distribution functions $\lambda(x;\beta,\gamma) = 1-(1+(x^2/c^2)^{\beta})^{-\gamma}$, where $\beta > 0$, $\gamma > 0$ and c is a scale parameter, which corresponds to $\eta_{\alpha/2}(T)$ in (5). A key difference between the two approaches is that while in Magnus's case c is a free parameter which should be optimised 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 the previous remark).

Remark 5 (Relationship with forecast combination) - Estimator (5) can be seen as a way to combine the two different forecasts $\hat{\theta}_T$ and $\tilde{\theta}$. Relative to the traditional methods on forecast combination (see, for instance, Granger and Newbold, 1986, chapter 9), our forecasting method has at least two advantages. First, traditional methods for forecast combinations require the availability of a series of out of sample forecasts for their implementation. This may be particularly problematic with short time series (an extreme case when traditional methods are not applicable is given by the example considered in section 4.1). Second, traditional methods do not account for the confidence level α associated to the subjective guess $\tilde{\theta}$, and therefore do not fully incorporate the judgmental elements of the decision making process. The method we propose can be directly applied to forecast combination procedures, by just testing the null hypothesis that the coefficient of $\tilde{\theta}$ is 1 and that of the alternative forecasts 0.

4 Risk Analysis

We evaluate the performance of the estimator proposed in the previous section with two simulations. They illustrate under which conditions the proposed estimator outperforms the classical and pretest estimator. The performance of the new estimator crucially depends on the quality of the subjective guess and the confidence associated to it.

4.1 Estimation of the mean of a univariate normal distribution with known variance

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 has been 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 (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 subjective guess respectively.

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

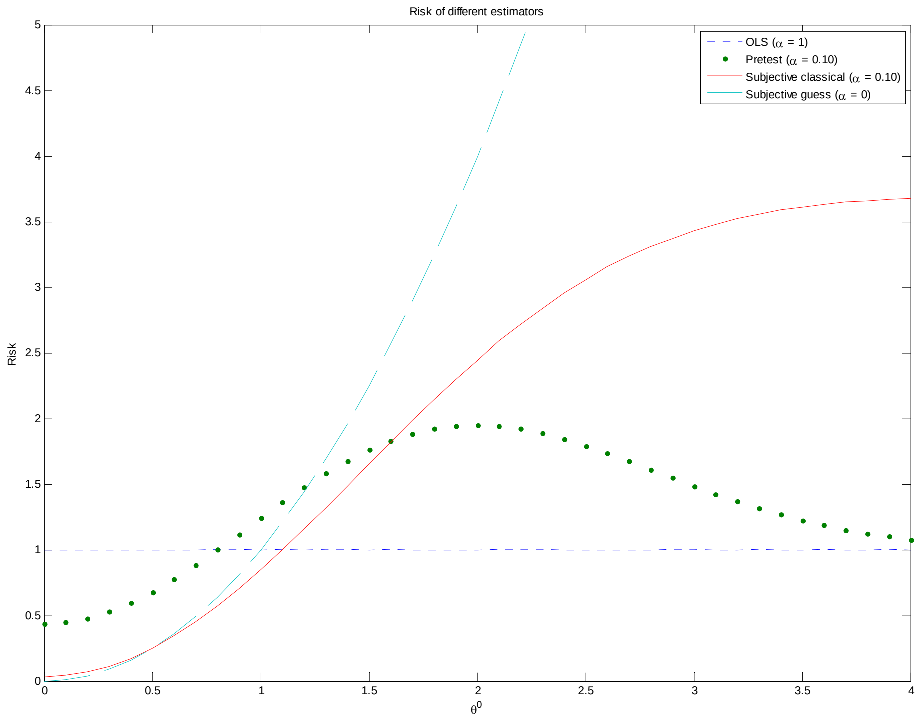
$$R(\theta^0; f(y)) = E_{\theta^0}[(f(y) - \theta^0)^2]$$

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 the OLS and the pretest have higher risk than the subjective classical estimators 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 subjective guess coincide. Furthermore, the new estimator has bounded risk (unless $\alpha = 0$, which corresponds to the subjective guess) and has always lower risk than the subjective guess, except for values of $\tilde{\theta}$ very close to θ^0. This is a very appealing feature of our estimator, as it guarantees that it will outperform any given subjective guess 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 subjective guess (i.e. the lower the α), the steeper its risk function, converging to the risk of the subjective guess as $\alpha \rightarrow 0$. The lower the confidence (i.e. the higher the α), the flatter its risk function, converging to the risk of OLS as $\alpha \rightarrow 1$.

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 subjective guess $\tilde{\theta}$ 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.

Figure 1 - Risk comparison of different estimators.



4.2 Monte Carlo Experiment

The relationship between precision of the subjective guess and performance of the estimator can be further illustrated by the following example. We generated random draws from a standard normal distribution, with different sample sizes, $T = 5, 20, 60, 120, 240, 1000$. For each series, we estimated the classical forecast estimator ($\hat{\theta}_T = T^{-1} \sum_{t=1}^{T} y_t$) and the one proposed in (5), θ_T^*, using a quadratic utility function, i.e. $\hat{U}_T(\theta) = -T^{-1} \sum_{t=1}^{T} (y_t - \theta)^2$. In the estimation of θ_T^*, we set $\alpha = 0.10$. Next, we computed the expected utility associated to these estimators and to the subjective guess $\tilde{\theta}$ ($E^i[U(\hat{\theta}_T)]$, $E^i[U(\theta_T^*)]$ and $E^i[U(\tilde{\theta})]$) with a Monte Carlo simulation (with $10,000$ random draws from the normal distribution). We repeated this procedure 5000 times and then averaged the

		T	**5**	**20**	**60**	**120**	**240**	**1000**
$\tilde{\theta}$	$E[U(\tilde{\theta})]$	$E[U(\hat{\theta}_T)]$	1.2068	1.0497	1.0167	1.0083	1.0039	1.0009
0	1		**1.0045**	**1.0012**	**1.0000**	**1.0003**	**0.9999**	**0.9999**
0.05	1.0025		**1.0071**	**1.0035**	**1.0022**	**1.0024**	**1.0020**	1.0016
0.1	1.01	$E[U(\theta_T^*)]$	**1.0142**	**1.0102**	**1.0085**	**1.0082**	1.0070	1.0034
0.5	1.2500		1.2151	1.1452	1.0621	1.0317	1.0150	1.0036
1	2.0000		1.6572	1.2019	1.0629	1.0317	1.0150	1.0036

Table 1: *Monte Carlo comparison of expected cost functions associated to the classical ($\hat{\theta}_T$) and subjective classical (θ_T^*) estimators for a given subjective guess $\tilde{\theta}$ and $\alpha = 0.10$. We formatted in bold the cases where the new estimator outperforms the classical estimator.*

expected utilities, i.e. $E[U(\hat{\theta}_T)] = \sum_{i=1}^{5000} E^i[U(\hat{\theta}_T)]/5000$ and the same for the other estimators. The results are reported in table 1.

The major points to be highlighted are the following. First, the new estimator θ_T^* may be biased but is consistent. Second, in small samples the classical estimator $\hat{\theta}_T$ performs worse than θ_T^* when the subjective guess $\tilde{\theta}$ is reasonably close to the true value. Third, in large samples, the performance of $\hat{\theta}_T$ and θ_T^* becomes roughly equivalent, independently of the subjective guess $\tilde{\theta}$.

These results have implications for the organisation of the forecasting process of any institution interested in forecasting. There should be a clear separation between the decision-maker providing the subjective guess and the confidence associated to it, and the econometrician whose task is to check whether such a subjective guess is supported by the available data or whether it can be improved. In particular, the formulation of the subjective guess should be independent of the data and econometric model used to evaluate it. In the previous example a subjective guess based on the OLS estimator would never be rejected by the data, but it would also have no value added. The responsibility of good or bad forecasts is therefore shared between the decision-maker and the econometrician. High confidence in a bad subjective guess would inevitably result in poor forecasts (in small samples). Therefore, formulating a good subjective guess may be as important as having a good econometric model.

5 Incorporating judgment into Classical Estimation

In this section we generalise the analysis of the previous sections. We formally define a new estimator which depends on a subjective guess 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 maximise 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 $\hat{U}_T(\theta)$ the finite sample approximation of the expected utility, which depends on the decision variables θ, data and sample size. θ is a vector belonging to the k-dimensional parameter space Θ. We assume the following:

Condition 1 (Uniform Convergence) *$\hat{U}_T(\theta)$ converges uniformly in probability to $U_0(\theta)$.*

Condition 2 (Identification) *$U_0(\theta)$ is uniquely maximised at θ^0.*

Condition 3 (Compactness) *Θ is compact.*

Condition 4 (Continuity) *$U_0(\theta)$ 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 maximises the empirical expected utility:

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

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

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

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

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

Condition 8 (Convexity) $U_0(\theta)$ *is globally convex in* θ.

Convexity of the objective function in θ is needed to ensure consistency, as otherwise the estimator could stop at a local maximum. The others are technical conditions typically imposed to derive asymptotic normality results.

Define the following shrinkage estimator, which shrinks from the subjective guess $\tilde{\theta}$ to the classical estimator $\hat{\theta}_T$:

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

Denote with $\hat{\Sigma}_T$ a $\sqrt{T}$-consistent estimate of Σ and let $\eta_{\alpha,k}$ denote the χ_k^2 critical value associated to the confidence level α. Define:[1]

$$\hat{z}_T(\theta_T^*(\lambda)) \equiv T \; \nabla_\theta' \hat{U}_T(\theta_T^*(\lambda)) \hat{\Sigma}_T^{-1} \nabla_\theta \hat{U}_T(\theta_T^*(\lambda)) \tag{7}$$

Notice that for each $\lambda \in [0,1]$ under the null hypothesis $H_0 : \theta_T^*(\lambda) = \theta^0$ we have:

$$\hat{z}_T(\theta_T^*(\lambda)) \xrightarrow{d} \chi_k^2$$

The new estimators is defined as follows:

[1] 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.

Definition 2 (Subjective Classical Estimator) *Let $\tilde{\theta}$ denote the subjective guess and $\alpha \in [0,1]$ the confidence level associated to it. Define $\theta_T^*(\lambda)$ and $\hat{z}_T(\theta_T^*(\lambda))$ as in (6) and (7), respectively. The* ***subjective classical estimator*** *is $\theta_T^*(\hat{\lambda})$, where:*

1. *if $\hat{z}_T(\theta_T^*(0)) \leq \eta_{\alpha,k}$, $\hat{\lambda} = 0$;*
2. *if $\hat{z}_T(\theta_T^*(0)) > \eta_{\alpha,k}$, $\hat{\lambda}$ is the solution to the following constrained maximisation problem:*

$$\begin{aligned} \max_{\lambda \in [0,1]} \quad & \hat{U}_T(\theta_T^*(\lambda)) \\ s.t. \quad & \hat{z}_T(\theta_T^*(\lambda)) = \eta_{\alpha,k} \end{aligned} \tag{8}$$

This estimator generalises estimator (5) proposed in section 3 to any objective function satisfying conditions $1-8$ and to any dimension of the decision variable θ. It first checks whether the given subjective guess $\tilde{\theta}$ is supported by the data. If for given $\tilde{\theta}$ and confidence level α, the objective function cannot be increased in a statistically significant way, the subjective guess $\tilde{\theta}$ is retained as the forecast estimator. Rejection of the null hypothesis implies that it is possible to move away from $\tilde{\theta}$ and increase (in a statistical sense) the objective function. The classical estimator $\hat{\theta}_T$ (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 subjective guess $\tilde{\theta}$ towards the classical estimator $\hat{\theta}_T$. The amount of shrinkage is determined by the constraint in (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.

The convexity assumption guarantees that the increase in the objective function is statistically significant $\forall \lambda \in [0, \hat{\lambda}]$, i.e. for every point along the path connecting $\tilde{\theta}$ to $\theta_T^*(\hat{\lambda})$. Notice that $\forall \lambda \in (\hat{\lambda}, 1]$ the probability of reducing instead of increasing the population value of the objective function (i.e., the probability that the first derivative in population switches sign) becomes greater than α. Therefore, by moving beyond $\theta_T^*(\hat{\lambda})$, and in particular by choosing the classical estimator $\hat{\theta}_T = \theta_T^*(1)$, the forecaster risks reducing, rather than increasing, her expected utility.

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 new estimator $\theta_T^*(\hat{\lambda})$ of Definition 2 satisfies the following properties:*

1. *If $\alpha = 1$, $\theta_T^*(\hat{\lambda})$ is the classical estimator;*
2. *If $\alpha > 0$, $\theta_T^*(\hat{\lambda})$ converges in probability to the classical estimator.*

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 $\pi(\theta)$. This prior distribution is then updated with the information contained in the sample, by applying the Bayes rule. The posterior density of

θ given the sample data $y^T \equiv \{y_t\}_{t=1}^T$ is given by:

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

where $f(y^T|\theta)$ denotes the sampling distribution and $m(y^T)$ the marginal distribution of y^T. In the Bayesian framework, non-sample information is incorporated in the econometric analysis via the prior $\pi(\theta)$. Once the prior has been formulated, Bayesian techniques can be applied to find the $\hat{\theta}_T^B$ which maximises the expected utility.[2]

The estimator proposed in this paper offers a classical alternative to Bayesian techniques to account for non-sample information in forecasting. The non-sample information is summarised by the two parameters $(\tilde{\theta}, \alpha)$. $\tilde{\theta}$ is the decision-maker's subjective guess about the optimal decision variables, while α is the confidence level in such a guess, which is then used to test the hypothesis $H_0 : \tilde{\theta} = \theta^0$, where θ^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 true parameter value θ^0 of the model. In this case, the prior distribution collapses to a degenerate distribution with total mass on the point θ^0. With such a prior, the posterior will be always identical to the prior, no matter what the sample data looks like and the Bayesian estimator will be $\hat{\theta}_T^B = \theta^0$. In our setting, on the other hand, certainty about parameter values may be expressed by setting $\tilde{\theta} = \theta^0$ and $\alpha = 0$. When $\alpha = 0$ the null hypothesis can never be rejected and the new estimator becomes $\theta_T^*(\hat{\lambda}) = \theta^0$.

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, lack of non-sample information can be

[2] Bayesian decision theory typically refers to loss function minimisation and distinguishes between model parameters and actions. See Berger (1985) and the references cited in Granger and Machina (2005) for further discussion.

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 (see, for instance, proposition 12.1 of Hamilton 1994). 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 $(\tilde{\theta}, \alpha)$. The choice between the two estimators depends on how the non-sample information is formalised. Our estimator is not suited to exploit non-sample 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 non-sample information is expressed in terms of the two parameters $(\tilde{\theta}, \alpha)$, 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 s/he provides the non-sample information.

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 maximising a mean-variance utility function. We highlight how the theory proposed in this paper naturally takes into account the impact of estimation errors and show with an out of sample exercise how the subjective classical estimator outperforms standard mean-variance optimisers.

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

6.1 Mean-Variance Asset Allocation

Markowitz's (1952) mean-variance model provides the standard benchmark for portfolio allocation. It formalises the intuition that investors optimise 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 colourfully, 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 maximise the empirical expected utility, without taking into consideration whether this maximisation is statistically significant or not. Our theory provides a natural alternative. For a given benchmark portfolio (the subjective guess $\tilde{\theta}$ 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.

To formalise this discussion, consider a portfolio with $N + 1$ assets. Denote with θ the N-vector of weights associated to the first N assets entering a given portfolio, and denote with $y_t(\theta)$ 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_{N+1} is the weight associated to the $(N + 1)^{th}$ asset of the portfolio. Let's assume an investor wants to maximise a trade-off between mean and variance of portfolio returns, resulting in the following expected utility function:

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

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

$$\hat{U}_T(\theta) = T^{-1}\sum_{t=1}^{T} y_t(\theta) - \lambda\{T^{-1}\sum_{t=1}^{T} y_t^2(\theta) - [T^{-1}\sum_{t=1}^{T} y_t(\theta)]^2\} \tag{10}$$

The first order conditions are:

$$\begin{aligned} \nabla_\theta \hat{U}(\theta) &= T^{-1}\sum_{t=1}^{T} \nabla_\theta y_t(\theta) - \\ &\quad -\lambda\{T^{-1}2\sum_{t=1}^{T} y_t(\theta)\nabla_\theta y_t(\theta) - 2[T^{-1}\sum_{t=1}^{T} y_t(\theta)]T^{-1}\sum_{t=1}^{T} \nabla_\theta y_t(\theta)\} \end{aligned}$$

where $\nabla_\theta y_t(\theta) \equiv y_t^N - y_t^{N+1}\iota$, 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 ι 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 $\lambda = 1$ and use as subjective guess the equally weighted portfolio and confidence levels $\alpha = 1, 0.10, 0.01, 0.001$.

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 (2007), 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 from $t = M$, we estimate the optimal weights using the previous M observations. Next, we compute the out of sample realised 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 realised utilities associated to the optimal portfolio out of sample returns. In table 2 we report the difference in average utilities between the optimal portfolio and the equal weight benchmark. That is, if we denote with $U_t(\theta) \equiv y_t(\theta) - \lambda\{y_t^2(\theta) - [(T-M)^{-1} \sum_{t=M+1}^{T} y_t(\theta)] y_t(\theta)\}$ the realised utility associated to portfolio θ, we compute:

$$Z_{T-M} \equiv (T-M)^{-1} \sum_{t=M+1}^{T} [U_t(\theta_{t-1}^{*}(\hat{\lambda})) - U_t(\tilde{\theta})] \tag{11}$$

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

	$M=60$			$M=120$		
	$N=4$	$N=16$	$N=30$	$N=4$	$N=16$	$N=30$
$\alpha=1$	-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)*	–

Table 2: *Average difference in out of sample realized utilities associated to $\theta_T^*(\hat{\lambda})$ 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 asteriscs, respectively.*

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 optimised portfolio outperforms the benchmark in a statistically significant way for $\alpha = 0.10$. When $\alpha = 0.01$, the difference in realised expected utilities is still positive but no longer statistically significant. For the portfolio with 16 assets, none of the optimised portfolios statistically outperforms the benchmark. For $N = 30$, the benchmark is not rejected at the 10% level.

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 realised 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 4.1, 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 4.1. The lower α, the higher the likelihood that the subjective classical estimator will have lower risk than the subjective guess (i.e., the risk function of the subjective classical estimator will lie below that of the subjective guess, except for values of θ^0 very close to $\tilde{\theta}$). At the same time, the lower α, the higher the risk associated to the subjective classical estimator for values of θ^0 far away from $\tilde{\theta}$.

6.2 Forecasting U.S. GDP

We illustrate how the theory of section 5 can be applied to forecast the U.S. real GDP. A possible difficulty in implementing the estimator is related to the formulation of a subjective guess on parameters of an econometric model about which the decision-maker may know nothing or very little. We propose a simple strategy to map a subjective guess on the variable of interest to the decision-maker (GDP in this case) into subjective guesses on the parameters of the econometrician's favourite model.

In principle, it is possible to express a subjective guess directly on the parameter vector θ or indirectly on the dependent variable y_{T+1} to be forecast. If the decision-maker can formulate a guess on θ, 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 behaviour of y_{T+1}, rather than about abstract model parameters. Let's denote this subjective guess as $\tilde{y}_{T+1}$. Using the notation of section 5, this can be translated into a subjective guess on θ as follows:

$$\begin{aligned} \tilde{\theta} &= \arg\max_{\theta} \hat{U}_T(\theta) \\ s.t.\ \hat{y}_{T+1}(\theta) &= \tilde{y}_{T+1} \end{aligned} \tag{12}$$

where $\hat{y}_{T+1}(\theta)$ is the model's forecast conditional on the parameter vector θ. The subjective guess $\tilde{y}_{T+1}$ is mapped into a subjective guess on the parameter vector by choosing the $\tilde{\theta}$ that maximises 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:

$$y_t = \theta_0 + \sum_{i=1}^{4} \theta_i y_{t-i} + \varepsilon_t \tag{13}$$

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

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

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

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

where $\hat{\theta}_T$ 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®

		$\tilde{y}_{T+1} = 3\%$		$\tilde{y}_{T+1} = 5\%$	
	$\hat{\theta}_T$	$\tilde{\theta}$	$\theta^*_T(\hat{\lambda})$	$\tilde{\theta}$	$\theta^*_T(\hat{\lambda})$
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%

Table 3: *Subjective guesses $\tilde{\theta}$ and estimated parameters $\theta^*_T(\hat{\lambda})$ 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 $\hat{\theta}_T$. A subjective guess of 5%, instead, is rejected by the data, resulting in parameter estimates different from the parameter guess.*

database[3]. 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 subjective guesses 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 3 show that $\tilde{y}_{T+1} = 3\%$ maps into a parameter guess $\tilde{\theta}$ which cannot be rejected by the data ($\theta^*_T(\hat{\lambda}) = \tilde{\theta}$). These parameter values are also very close to the OLS estimates $\hat{\theta}_T$, resulting in very similar forecasts. Note that in this case the forecast associated to $\theta^*_T(\hat{\lambda})$ is equal to 3%, the original subjective guess ($\tilde{y}_{T+1} = 3\%$).

The other subjective guess, $\tilde{y}_{T+1} = 5\%$, is instead rejected by the data at

[3]See http://research.stlouisfed.org/fred2.

the chosen confidence level, resulting in parameter estimates $\theta_T^*(\hat{\lambda})$ which are different from the parameter guess $\tilde{\theta}$. The estimated shrinkage factor $\hat{\lambda}$ was 0.68. The out of sample GDP forecast at Q4 2005 associated to $\theta_T^*(\hat{\lambda})$ is 3.77% and the OLS forecast is 3.19%, both definitely lower than the subjective guess of 5%.

7 Conclusion

Classical forecasts typically ignore non-sample information and estimation errors due to finite sample approximations. In this paper we pointed out how these two problems are connected. We argued that forecasts should optimise the objective function in a statistical sense, rather than in the usual deterministic way. We explicitly introduced into the classical estimation framework two elements: a subjective guess on the variable to be forecast and a confidence associated to it. Their role is to explicitly take into consideration the non-sample information available to the decision-maker. These elements served to define a new estimator, which statistically optimises the objective function, and to formalise the interaction between judgment and data in the forecasting process.

We provided three applications, which give strong support to our theory. We argued that there should be a clear separation between the decision-maker — who should provide the subjective guess and the confidence associated to it — and the econometrician — whose task is to check whether such subjective guess is supported by the available data or whether it can be improved. We illustrated how our new estimator may provide a satisfactory solution to the well-known implementation problems of the mean-variance asset allocation model. Finally, we showed how a subjective guess on the variable to be forecast can be mapped into a subjective guess on the parameters of the econometrician's favourite model.

8 Appendix

Proof of Theorem 1 (Properties of the New Estimator) - 1. If $\alpha = 1$, $\eta_{\alpha,k} = 0$ and the constraint in (8) becomes $\hat{z}_T(\theta^*_T(\lambda)) = 0$. This implies $\nabla_\theta \hat{U}_T(\theta^*_T(\hat{\lambda})) = 0$, which coupled with (8) implies $\theta^*_T(\hat{\lambda}) = \hat{\theta}_T$, where $\hat{\theta}_T$ is defined in Definition 1.

2. Let $\theta^0 \equiv \underset{T\to\infty}{p\lim}\hat{\theta}_T$. We need to show that $\theta^*_T(\hat{\lambda}) \xrightarrow{p} \theta^0$. By (8) and Condition 2, this is equivalent to show that $||\nabla_\theta \hat{U}_T(\theta^*_T(\hat{\lambda}))|| \xrightarrow{p} 0$. Note that definition 2 implies $\Pr(\hat{z}_T(\theta^*_T(\hat{\lambda})) > \eta_{\alpha,k}) = 0$. Suppose by contradiction that $||\nabla_\theta \hat{U}_T(\theta^*_T(\hat{\lambda}))|| \xrightarrow{p} c \neq 0$. Then $\Pr(\hat{z}_T(\theta^*_T(\hat{\lambda})) > \eta_{\alpha,k}) = \Pr(\nabla'_\theta \hat{U}_T(\theta^*_T(\hat{\lambda}))\hat{\Sigma}_T^{-1}\nabla_\theta \hat{U}_T(\theta^*_T(\hat{\lambda})) > \eta_{\alpha,k}/T)$. But since $\nabla'_\theta \hat{U}_T(\theta^*_T(\hat{\lambda}))\hat{\Sigma}_T^{-1}\nabla_\theta \hat{U}_T(\theta^*_T(\hat{\lambda}))$ 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 $T > T^*$, $\Pr(\nabla'_\theta \hat{U}_T(\theta^*_T(\hat{\lambda}))\hat{\Sigma}_T^{-1}\nabla_\theta \hat{U}_T(\theta^*_T(\hat{\lambda})) > \eta_{\alpha,k}/T) > q$. This implies a violation of the constraint in (8) and therefore a contradiction. ■

References

[1] **Berger, J.O. (1985)**, *Statistical Decision Theory and Bayesian Analysis*, 2^{nd} edition, Springer-Verlag, New York.

[2] **DeMiguel, V, L. Garlappi and R. Uppal (2007)**, Optimal versus Naive Diversification: How Inefficient is the $1/N$ Portfolio Strategy?, *Review of Financial Studies*, forthcoming.

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

[4] **Giacomini, R. and H. White (2006)**, Test of Conditional Predictive Ability, *Econometrica*, **74**: 1545-1578.

[5] **Granger, C. W. J. and M. Machina (2005)**, Forecasting and Decision Theory, forthcoming in the *Handbook of Economic Forecasting*, edited by G. Elliott, C. W. J. Granger and A. Timmermann.

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

[7] **Haavelmo, T. (1944)**, The Probability Approach in Econometrics, *Econometrica* **12**, supplement.

[8] **Hamilton, J.D. (1994)**, *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

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

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

[11] **Magnus, J.R. and J. Durbin (1999)**, Estimation of Regression Coefficients of Interest when Other Regression Coefficients Are of No Interest, *Econometrica*, **67**: 639-643

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

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

[14] **Newey, W. and D. McFadden (1994)**, Large Sample Estimation and Hypothesis Testing, in *Handbook of Econometrics*, Volume IV, edited by R.F. Engle and D.L. McFadden, Elsevier Science.

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