Journal of Forecasting
J. Forecast. **23**, 251–274 (2004)
Published online in Wiley InterScience (www.interscience.wiley.com). **DOI**: 10.1002/for.916

Forecast Accuracy After Pretesting with an Application to the Stock Market

DMITRY DANILOV[1] AND JAN R. MAGNUS[2]*
[1] *Eurandom, Eindhoven University of Technology, The Netherlands*
[2] *CentER, Tilburg University, The Netherlands*

ABSTRACT

In econometrics, as a rule, the same data set is used to select the model and, conditional on the selected model, to forecast. However, one typically reports the properties of the (conditional) forecast, ignoring the fact that its properties are affected by the model selection (pretesting). This is wrong, and in this paper we show that the error can be substantial. We obtain explicit expressions for this error. To illustrate the theory we consider a regression approach to stock market forecasting, and show that the standard predictions ignoring pretesting are much less robust than naive econometrics might suggest. We also propose a forecast procedure based on the 'neutral Laplace estimator', which leads to an improvement over standard model selection procedures.

KEY WORDS pretest; forecasting; model selection; stock returns

INTRODUCTION

In econometrics we typically use the same data for both model selection and forecasting (and estimation). Standard statistical theory is therefore not directly applicable, because the properties of forecasts (and estimates) depend not only on the stochastic nature of the selected model, but also on the way this model was selected.

The simplest example of this situation is the standard linear model $\boldsymbol{y} = \boldsymbol{X\beta} + \gamma\boldsymbol{z} + \boldsymbol{\varepsilon}$, where we are uncertain whether to include $\boldsymbol{z}$ or not. The usual procedure is to compute the t-statistic for γ, and then, depending on whether $|t|$ is 'large' or 'small', decide to use the unrestricted or the restricted (with $\gamma = 0$) model. We then forecast y_{n+1} from the selected model. This forecast is a *pretest* forecast, but we commonly report its properties as if forecasting had not been preceded by model selection. This is clearly wrong. We should correctly report the bias and variance (or mean squared error) of the forecasts, taking full account of the fact that model selection and forecasting are an integrated procedure. This paper attempts to do this, both in theory and practice.

* Correspondence to: Jan R. Magnus, CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. E-mail: magnus@uvt.nl

The theoretical contribution of the paper is the 'equivalence theorem for forecasting' (Theorem 1), where we present explicit expressions for the forecast bias and the mean squared forecast error, taking full account of the effects of pretesting.

In order to investigate the effects of ignoring pretesting on forecasts in practice, we reconsider a question from the finance literature, discussed by Pesaran and Timmermann (1994): can the annual excess returns on common stocks for the Standard & Poor 500 index be predicted? The application of linear regression in finance probably started with the capital asset pricing model. Black *et al*. (1972) proposed a linear regression model to explain empirically observed asset returns. Fama and MacBeth (1973) introduced a cross-section approach, and regressed the asset's excess return on the intercept and the β's of the CAPM model. Subsequent studies extended the set of explanatory variables, see in particular Rozeff (1984), French *et al*. (1987), Fama and French (1989), Chen *et al*. (1986), Balvers *et al*. (1990), Fama and French (1992) and Cheng *et al*. (1990). Pesaran and Timmermann (1994) demonstrated that a regression model preceded by model selection can actually predict movements of the Dow Jones and Standard & Poor 500 indexes with a sufficient degree of accuracy. This result was enriched and reinforced in Pesaran and Timmermann (1995), where a number of model selection criteria were employed. The problem of forecasting the market moves was reconsidered in Granger and Pesaran (2000), where the authors argue that not a point stock value but rather the probability of the fall in the stock market is the key element, and propose a way to estimate this probability.

We find that pretesting matters a lot. Both the statistical and the financial conclusions are much affected if pretesting is ignored. In addition, we ask whether an improvement over the standard (discontinuous) model selection procedures might be possible, and we show that this is indeed possible, based on so-called Laplace weights. Table I later suggests that model selection based on Laplace weights leads to better forecasts, both from the statistical and from the financial point of view.

In finance, pretesting is usually called 'data-snooping'. Lo and MacKinlay (1990) noted that tests of financial asset pricing models may yield misleading inferences when properties of the data are used to construct the test statistics, that is, when pretesting has taken place. Foster *et al*. (1997) worried about the distribution of R^2 after pretesting. White (2000) developed a reality check bootstrap methodology which provides a step towards answering the question whether the superior performance of the preferred model is due to superior economic content or to luck. Sullivan *et al*. (1999) showed that White's methodology can be applied in practice, and that pretesting matters.

Our results differ from those of White in several important respects. Apart from some differences in set-up, our Theorem 1 provides an (exact) finite-sample result and not an asymptotic result as in White. Secondly, we provide bounds for our pretest forecasts, thus allowing rather precise probability statements.

The paper is organized as follows. The next section contains the set-up and notation and reviews some earlier results, which are required for the development of the theory. The main result is presented in the third section (Theorem 1), giving the bias, variance and mean squared forecast error of the pretest forecast (in fact, of the WALS forecast, a generalization of the pretest forecast). In the fourth section we apply the theory to the problem of forecasting stock market moves and show that results ignoring pretesting are much less robust than naive econometrics would seem to imply. In the fifth section we present a continuous analogue of pretesting which can greatly improve the properties of forecasts. Some conclusions are offered in the final section.

SET-UP, NOTATION AND PRELIMINARY RESULTS

The set-up is the same as in Magnus and Durbin (1999) and Danilov and Magnus (2004), hereafter DM04. We consider the standard linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{y}$ $(n \times 1)$ is the vector of observations, $\boldsymbol{X}$ $(n \times k)$ and $\boldsymbol{Z}$ $(n \times m)$ are matrices of nonrandom regressors, $\boldsymbol{\varepsilon}$ $(n \times 1)$ is a random vector of unobservable disturbances, and $\boldsymbol{\beta}$ $(k \times 1)$ and $\boldsymbol{\gamma}$ $(m \times 1)$ are unknown nonrandom parameter vectors.[1] We assume that $k \geq 1$, $m \geq 1$, $n - k - m \geq 1$, that the design matrix $(\boldsymbol{X}:\boldsymbol{Z})$ has full column rank $k + m$, and that the disturbances $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are i.i.d. $\mathrm{N}(0, \sigma^2)$.[2]

The reason for distinguishing between $\boldsymbol{X}$ and $\boldsymbol{Z}$ is that $\boldsymbol{X}$ contains explanatory variables ('focus' regressors) that we want in the model on theoretical or other grounds, while $\boldsymbol{Z}$ contains additional explanatory variables ('auxiliary' regressors) of which we are less certain. We define the matrices

$$\boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' \quad \text{and} \quad \boldsymbol{Q} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z})^{-1/2}$$

and the normalized parameter vector $\boldsymbol{\theta} = (\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z})^{1/2}\boldsymbol{\gamma}$. The least-squares (LS) estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are $\boldsymbol{b}_u = \boldsymbol{b}_r - \boldsymbol{Q}\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\gamma}} = (\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z})^{-1/2}\hat{\boldsymbol{\theta}}$, where $\boldsymbol{b}_r = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ and $\hat{\boldsymbol{\theta}} = (\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z})^{-1/2}\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{y}$. The subscripts '$u$' and '$r$' denote 'unrestricted' and 'restricted' (with $\boldsymbol{\gamma} = \boldsymbol{0}$) respectively. Notice that $\hat{\boldsymbol{\theta}} \sim \mathrm{N}(\boldsymbol{\theta}, \sigma^2\boldsymbol{I}_m)$ and that $\boldsymbol{b}_r$ and $\hat{\boldsymbol{\theta}}$ are independently distributed.

Let $\boldsymbol{S}_i$ be an $m \times r_i$ selection matrix of rank r_i $(0 \leq r_i \leq m)$, so that $\boldsymbol{S}_i' = (\boldsymbol{I}_{r_i}:\boldsymbol{O})$ or a column permutation thereof. The equation $\boldsymbol{S}_i'\boldsymbol{\gamma} = \boldsymbol{0}$ thus selects a subset of the γ's to be equal to zero. Following DM04, the LS estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ under the restriction $\boldsymbol{S}_i'\boldsymbol{\gamma} = \boldsymbol{0}$ are then given by

$$\boldsymbol{b}_{(i)} = \boldsymbol{b}_r - \boldsymbol{Q}\boldsymbol{W}_i\hat{\boldsymbol{\theta}}, \qquad \boldsymbol{c}_{(i)} = (\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z})^{-1/2}\boldsymbol{W}_i\hat{\boldsymbol{\theta}}$$

where

$$\boldsymbol{W}_i = \boldsymbol{I}_m - (\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z})^{-1/2}\boldsymbol{S}_i\left(\boldsymbol{S}_i'(\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z})^{-1}\boldsymbol{S}_i\right)^{-1}\boldsymbol{S}_i'(\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z})^{-1/2}$$

is a symmetric idempotent $m \times m$ matrix of rank $m - r_i$. (If $r_i = 0$ then $\boldsymbol{W}_i = \boldsymbol{I}_m$.) The distribution of $\boldsymbol{b}_{(i)}$ is given by

$$\boldsymbol{b}_{(i)} \sim \mathrm{N}\left(\boldsymbol{\beta} + \boldsymbol{Q}(\boldsymbol{I}_m - \boldsymbol{W}_i)\boldsymbol{\theta}, \sigma^2\left((\boldsymbol{X}'\boldsymbol{X})^{-1} + \boldsymbol{Q}\boldsymbol{W}_i\boldsymbol{Q}'\right)\right) \tag{2}$$

There are 2^m different models to consider, one for each subset of $\gamma_1, \ldots, \gamma_m$ set equal to zero. A *pretest* estimator of $\boldsymbol{\beta}$ is obtained by first selecting one of these models (using t- or F-tests or other model selection criteria), and then estimating $\boldsymbol{\beta}$ in the selected model. We shall assume throughout that the model selection is based exclusively on the residuals from the restricted model, that is, on

[1] We follow the notation proposed in Abadir and Magnus (2002).

[2] In contrast to our estimation paper, we may allow $k = 0$ here, in which case $\boldsymbol{X}$ is absent. All subsequent results hold in that case, but some care needs to be taken about the interpretation of the formulas.

$\boldsymbol{My}$. This assumption appears to be satisfied in all standard cases. (Note that the residuals in the ith model can always be expressed as $\boldsymbol{e}_{(i)} = \boldsymbol{D}_i\boldsymbol{My}$ for some idempotent matrix $\boldsymbol{D}_i$; see DM04, lemma A1.) More generally, a WALS (weighted-average least-squares) estimator of $\boldsymbol{\beta}$ is defined as $\boldsymbol{b} = \Sigma_i\lambda_i\boldsymbol{b}_{(i)}$, where the weights satisfy

$$\lambda_i = \lambda_i(\boldsymbol{My}), \qquad \lambda_i \geq 0, \qquad \sum_i \lambda_i = 1$$

and the sum is taken over all 2^m models. The pretest estimator is a special case of the WALS estimator when all λ_i are 0 except one which is 1. The standard model selection procedures (such as general-to-specific or specific-to-general) all fall in this class, so that they can be captured by an appropriate choice of the λ_i.

The WALS estimator of $\boldsymbol{\beta}$ can be written as $\boldsymbol{b} = \boldsymbol{b}_r - \boldsymbol{QW}\hat{\boldsymbol{\theta}}$, where $\boldsymbol{W} = \Sigma_i\lambda_i\boldsymbol{W}_i$. Notice that $\boldsymbol{W}$ is a random matrix, because the $\{\lambda_i\}$ are random. The equivalence theorem for estimation (DM04, theorem 1) now says that

$$\mathrm{E}(\boldsymbol{b}) = \boldsymbol{\beta} - \boldsymbol{Q}\mathrm{E}(\boldsymbol{W}\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}), \qquad \mathrm{var}(\boldsymbol{b}) = \sigma^2(\boldsymbol{X'X})^{-1} + \boldsymbol{Q}\,\mathrm{var}(\boldsymbol{W}\hat{\boldsymbol{\theta}})\boldsymbol{Q'}$$

and hence that

$$\mathrm{MSE}(\boldsymbol{b}) = \sigma^2(\boldsymbol{X'X})^{-1} + \boldsymbol{Q}\mathrm{MSE}(\boldsymbol{W}\hat{\boldsymbol{\theta}})\boldsymbol{Q'}$$

showing that the properties of the complicated WALS (pretest) estimator $\boldsymbol{b}$ of $\boldsymbol{\beta}$ depend critically on the properties of the less complicated estimator $\boldsymbol{W}\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. The distribution of $\boldsymbol{W}\hat{\boldsymbol{\theta}}$ thus plays a central role, both in the current paper and in our estimation paper, where the properties of $\boldsymbol{W}\hat{\boldsymbol{\theta}}$ are discussed in detail for the special case $\sigma^2 = 1$. For example, figure 1 in DM04 presents the bias, variance and mean squared error of $\boldsymbol{W}\hat{\boldsymbol{\theta}}$ in the case $m = 1$, and section 6 provides an extensive discussion for the case $m = 2$. It turns out that $\mathrm{MSE}(\boldsymbol{W}\hat{\boldsymbol{\theta}})$ can be unbounded in some cases, in particular, in the specific-to-general procedure. However, if we 'orthogonalize' the auxiliary regressors such that $\boldsymbol{Z'MZ} = \boldsymbol{I}_m$, then $\mathrm{MSE}(\boldsymbol{W}\hat{\boldsymbol{\theta}})$ remains bounded; this 'orthogonalization' has other advantages as well.

THE EQUIVALENCE THEOREM FOR FORECASTING

Suppose now that our interest is in forecasting rather than estimation. We assume that the data are generated by (1), possibly with one or more of the γ_i equal to zero. Under the restriction $\boldsymbol{S}_i'\boldsymbol{\gamma} = \boldsymbol{0}$, the one-period-ahead LS forecast is given by

$$\begin{aligned}\hat{y}_{n+1}^{(i)} &= \boldsymbol{x}_{n+1}'\boldsymbol{b}_{(i)} + \boldsymbol{z}_{n+1}'\boldsymbol{c}_{(i)} \\ &= \boldsymbol{x}_{n+1}'(\boldsymbol{b}_r - \boldsymbol{QW}_i\hat{\boldsymbol{\theta}}) + \boldsymbol{z}_{n+1}'\left((\boldsymbol{Z'MZ})^{-1/2}\boldsymbol{W}_i\hat{\boldsymbol{\theta}}\right) \\ &= \boldsymbol{x}_{n+1}'\boldsymbol{b}_r - \boldsymbol{\zeta}_{n+1}'\boldsymbol{W}_i\hat{\boldsymbol{\theta}}\end{aligned}$$

where

$$\boldsymbol{\zeta}_{n+1} = \boldsymbol{Q'x}_{n+1} - (\boldsymbol{Z'MZ})^{-1/2}\boldsymbol{z}_{n+1}$$

and $\boldsymbol{x}_{n+1}$ and $\boldsymbol{z}_{n+1}$ denote next period's values of the focus and auxiliary regressors, respectively. Since the actual choice of model is uncertain and depends on the data and the model selection procedure, the forecast could be based on any of the 2^m available models (or a linear combination thereof). Hence the WALS forecast takes the form

$$\hat{y}_{n+1} = \sum_i \lambda_i \hat{y}_{n+1}^{(i)} = \boldsymbol{x}'_{n+1}\boldsymbol{b}_r - \boldsymbol{\zeta}'_{n+1}\boldsymbol{W}\hat{\boldsymbol{\theta}} \tag{3}$$

Since $y_{n+1} = \boldsymbol{x}'_{n+1}\boldsymbol{\beta} + \boldsymbol{z}'_{n+1}\boldsymbol{\gamma} + \varepsilon_{n+1}$, we obtain the forecast error (FE) as

$$\begin{aligned}\text{FE} &= \hat{y}_{n+1} - y_{n+1} \\ &= \boldsymbol{x}'_{n+1}(\boldsymbol{b}_r - \boldsymbol{\beta}) - \boldsymbol{\zeta}'_{n+1}\boldsymbol{W}\hat{\boldsymbol{\theta}} - \boldsymbol{z}'_{n+1}(\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z})^{-1/2}\boldsymbol{\theta} - \varepsilon_{n+1} \\ &= \boldsymbol{x}'_{n+1}(\boldsymbol{b}_r - \boldsymbol{\beta} - \boldsymbol{Q}\boldsymbol{\theta}) - \boldsymbol{\zeta}'_{n+1}(\boldsymbol{W}\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) - \varepsilon_{n+1}\end{aligned}$$

The following properties of the forecast error can now be established.

Theorem 1 (Equivalence theorem for forecasting) The WALS forecast error has the following expectation, variance and mean squared error:

$$\text{E}(\text{FE}) = -\boldsymbol{\zeta}'_{n+1}\text{E}(\boldsymbol{W}\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$$

$$\text{var}(\text{FE}) = \sigma^2\left(\boldsymbol{x}'_{n+1}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_{n+1} + 1\right) + \boldsymbol{\zeta}'_{n+1}\,\text{var}(\boldsymbol{W}\hat{\boldsymbol{\theta}})\boldsymbol{\zeta}_{n+1}$$

$$\text{MSFE} = \sigma^2\left(\boldsymbol{x}'_{n+1}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_{n+1} + 1\right) + \boldsymbol{\zeta}'_{n+1}\text{MSE}(\boldsymbol{W}\hat{\boldsymbol{\theta}})\boldsymbol{\zeta}_{n+1}$$

Proof: The essential ingredient is that $\boldsymbol{b}_r$ and $\boldsymbol{M}\boldsymbol{y}$ are independent, because they are jointly normal and uncorrelated since $\boldsymbol{M}\boldsymbol{X} = \boldsymbol{O}$. This implies that $\boldsymbol{b}_r$ and $\boldsymbol{W}\hat{\boldsymbol{\theta}}$ are independent, and hence that ($\boldsymbol{b}_r$, $\boldsymbol{W}\hat{\boldsymbol{\theta}}$, ε_{n+1}) are all independent of each other. The results follow. □

The importance of Theorem 1 is twofold. First, it gives explicit expressions for the first two moments of the forecast error, where we notice that these moments depend on $\boldsymbol{\theta}$ and σ^2, but not on $\boldsymbol{\beta}$. Second, it helps us to find an *optimal* forecast, in the sense of minimizing the mean squared forecast error. If we can find λ_i's such that $\boldsymbol{W}\hat{\boldsymbol{\theta}}$ is an optimal estimator of $\boldsymbol{\theta}$ then *the same* λ_i's will provide an optimal forecast. (These λ_i's are also the ones which provide the optimal WALS estimator of $\boldsymbol{\beta}$.) The question of finding an optimal estimator of $\boldsymbol{\theta}$ was studied in Magnus (2002), and led to the 'neutral Laplace' estimator. In a later section we shall apply the Laplace weights to forecasting, and investigate the usefulness of this approach.

Theorem 1 thus gives the actual (true) moments of the forecast error, taking into account that pretesting has occurred. In a typical applied paper, however, one does not take pretesting into account. Consequently, the bias of the forecast (that is, the expectation of the forecast error) is reported to be zero, and the reported MSFE (variance), denoted $\widetilde{\text{MSFE}}$, is given by

$$\widetilde{\text{MSFE}} = s^2\left(\boldsymbol{x}'_{n+1}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_{n+1} + \boldsymbol{\zeta}'_{n+1}\boldsymbol{W}\boldsymbol{\zeta}_{n+1} + 1\right) \tag{4}$$

where s^2 denotes the pretest estimator of σ^2, that is

$$s^2 = \sum_i \lambda_i s^2_{(i)}, \qquad s^2_{(i)} = \frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}_{(i)} - \boldsymbol{Z}\boldsymbol{c}_{(i)})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}_{(i)} - \boldsymbol{Z}\boldsymbol{c}_{(i)})}{n - k - m + r_i}$$

The reported MSFE is a random variable, not only because s^2 is random but also because $\boldsymbol{W}$ is random. The matrix $\boldsymbol{W}$ takes the value $\boldsymbol{W}_i$ and s^2 takes the value $s^2_{(i)}$, both with probability $\mathrm{E}(\lambda_i)$. The reported 95% prediction interval for y_{n+1} is

$$\hat{y}_{n+1} \pm 1.96 s \sqrt{\boldsymbol{x}'_{n+1}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_{n+1} + \boldsymbol{\zeta}'_{n+1}\boldsymbol{W}\boldsymbol{\zeta}_{n+1} + 1} \tag{5}$$

where the number 1.96 can of course be adjusted to take into account the degrees of freedom.

In contrast, if we take proper account of the effects of model selection, then the actual value of the forecast $\hat{y}_{n+1}$ remains the same, but its moments are quite different. Let us define the two functions

$$\psi_1(\boldsymbol{\theta}) := \boldsymbol{\zeta}'_{n+1}\mathrm{E}(\boldsymbol{W}\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \tag{6}$$

and

$$\psi_2(\boldsymbol{\theta}) := \sigma^2 \boldsymbol{x}'_{n+1}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_{n+1} + \boldsymbol{\zeta}'_{n+1}\mathrm{var}(\boldsymbol{W}\hat{\boldsymbol{\theta}})\boldsymbol{\zeta}_{n+1} \tag{7}$$

Then, by Theorem 1

$$\mathrm{FE} \sim (-\psi_1(\boldsymbol{\theta}), \psi_2(\boldsymbol{\theta}) + \sigma^2) \tag{8}$$

We consider two implications of this result, both of which will be employed in the next section. First, since $\mathrm{E}(\hat{y}_{n+1} + \psi_1(\boldsymbol{\theta})) = \mathrm{E}(y_{n+1})$, we define the (theoretical) 'bias-corrected' forecast as $\hat{y}_{n+1} + \psi_1(\boldsymbol{\theta})$, and obtain an estimated bias-corrected forecast by replacing $\boldsymbol{\theta}$ with $\hat{\boldsymbol{\theta}}$. When the number of observations n becomes large, then $\hat{\boldsymbol{\theta}}$ will not converge to $\boldsymbol{\theta}$, because $\mathrm{var}(\hat{\boldsymbol{\theta}}) = \sigma^2 \boldsymbol{I}_m$. Hence, $\hat{\boldsymbol{\theta}}$ is an unbiased but not a consistent estimator of $\boldsymbol{\theta}$. To protect ourselves against 'large' deviations of $\hat{\boldsymbol{\theta}}$ from $\boldsymbol{\theta}$, we shall also consider the conservative interval around the estimated bias-corrected forecast:

$$\hat{y}_{n+1} + \underline{C}_1(\hat{\boldsymbol{\theta}}, \hat{\sigma}) < \hat{y}_{n+1} + \psi_1(\boldsymbol{\theta}) < \hat{y}_{n+1} + \overline{C}_1(\hat{\boldsymbol{\theta}}, \hat{\sigma}) \tag{9}$$

where

$$\underline{C}_1(\hat{\boldsymbol{\theta}}, \sigma) := \min_{\boldsymbol{\theta} \in \mathcal{H}(\hat{\boldsymbol{\theta}})} \psi_1(\boldsymbol{\theta}), \qquad \overline{C}_1(\hat{\boldsymbol{\theta}}, \sigma) := \max_{\boldsymbol{\theta} \in \mathcal{H}(\hat{\boldsymbol{\theta}})} \psi_1(\boldsymbol{\theta})$$

$\hat{\sigma}$ is a consistent estimator of σ, the set $\mathcal{H}$ is an m-dimensional cube, defined by $\mathcal{H}(\hat{\boldsymbol{\theta}}) := \{\boldsymbol{\theta} : |\hat{\theta}_i - \theta_i| < a_m \sigma,\ i = 1, \ldots, m\}$, and a_m is determined such that, for standard normal u, $(\Pr(|u| < a_m))^m = 0.95$. In our application, $m = 4$ so that $a_m = 2.49$.

Second, an approximate (theoretical) 95% prediction interval for y_{n+1} is given by

$$\hat{y}_{n+1} + \psi_1(\boldsymbol{\theta}) \pm 1.96 \sqrt{\psi_2(\boldsymbol{\theta}) + \sigma^2} \tag{10}$$

The interval is approximate because the distribution of FE is not normal. Furthermore, in contrast to (5), the interval depends on $\boldsymbol{\theta}$ (and on σ), which is unknown. We obtain an estimated prediction interval by replacing $\boldsymbol{\theta}$ with $\hat{\boldsymbol{\theta}}$ and σ with $\hat{\sigma}$. Again, we protect ourselves against 'large' deviations of $\hat{\boldsymbol{\theta}}$ from $\boldsymbol{\theta}$ by also considering the more conservative interval

$$\hat{y}_{n+1} + \underline{C}_2(\hat{\boldsymbol{\theta}}, \hat{\sigma}) < y_{n+1} < \hat{y}_{n+1} + \overline{C}_2(\hat{\boldsymbol{\theta}}, \hat{\sigma}) \tag{11}$$

where

$$\underline{C}_2(\hat{\boldsymbol{\theta}}, \sigma) := \min_{\boldsymbol{\theta} \in \mathcal{H}(\hat{\boldsymbol{\theta}})} \left(\psi_1(\boldsymbol{\theta}) - 1.96\sqrt{\psi_2(\boldsymbol{\theta}) + \sigma^2}\right)$$

and

$$\overline{C}_2(\hat{\boldsymbol{\theta}}, \sigma) := \max_{\boldsymbol{\theta} \in \mathcal{H}(\hat{\boldsymbol{\theta}})} \left(\psi_1(\boldsymbol{\theta}) + 1.96\sqrt{\psi_2(\boldsymbol{\theta}) + \sigma^2}\right)$$

Later we shall also need (conservative) prediction intervals for $\boldsymbol{x}'_{n+1}\boldsymbol{\beta} + \boldsymbol{z}'_{n+1}\boldsymbol{\gamma}(\mathrm{E}y_{n+1})$ and for $\psi_2(\boldsymbol{\theta}) + \psi_1^2(\boldsymbol{\theta}) + \sigma^2$ (the MSFE).

FORECASTING STOCK RETURNS

In order to investigate the effects of ignoring pretesting on forecasts in practice, we consider a question from the finance literature. We shall reconsider the question discussed by Pesaran and Timmermann (1994), hereafter PT94, and others: can the annual excess returns on common stocks for the Standard & Poor 500 (SP 500) index be predicted? Our analysis is not meant as a criticism of PT94. Since we do not have exactly the same data set, and our questions are different in some details, we shall compare *our own analysis* when pretesting is not taken into account with *our own analysis* when pretesting is taken into account.

The dependent variable in the linear regression is y_t, the excess returns in year t. In analysing the effect of pretesting we have to decide which regressors play a role, and which of these are focus regressors and which are auxiliary. We selected four focus regressors ($k = 4$) and four auxiliary regressors ($m = 4$). The focus regressors are:

constant term
PI_{t-2}: annual inflation rate (lagged two periods)
$\mathrm{DI3}_{t-1}$: change in 3-month T-bill rate (lagged one period)
SPREAD_{t-1}: credit spread (lagged one period)

The auxiliary regressors are:

YSP_{t-1}: dividend yield on SP 500 portfolio (lagged one period)
DIP_{t-1}: annual change in industrial production (lagged one period)
PER_{t-1}: price–earnings ratio (lagged one period)
DLEAD_{t-2}: annual change in leading business cycle indicator (lagged two periods)

We could not acquire *exactly* the same data set as PT94, but we almost could. A full description and all the data are given in an appendix. Our data set contains eight annual time series (plus a constant term) over 46 years (1956–2001).

Employing a forward (specific-to-general) model selection procedure, only one of the four auxiliary regressors is selected, namely YSP_{t-1}, and we obtain the following estimated model of the annual excess returns over the period 1956–1991:

$$\hat{y}_t = \underset{(0.084)}{-0.343} - \underset{(0.44)}{1.65}\,\mathrm{PI}_{t-2} - \underset{(0.02)}{0.04}\,\mathrm{DI3}_{t-1} + \underset{(0.04)}{0.17}\,\mathrm{SPREAD}_{t-1} + \underset{(2.17)}{10.14}\,\mathrm{YSP}_{t-1}$$

with $R^2 = 0.655$, $\overline{R}^2 = 0.611$ and DW = 2.54. The selected model is the same as the model selected in PT94 (p. 339), estimated over the period 1954–1991.

A few words of explanation are in order. First, the *forward* pretest procedure (also called specific-to-general) is defined by starting from the smallest model (the restricted model) with k explanatory variables (the $\boldsymbol{X}$-variables). We first estimate the m models with one additional regressor. If none of the m t-statistics is significant, we choose the restricted model. If at least one of the t-statistics is significant, we select the regressor whose t-statistic is the largest (in absolute value), and keep this regressor in the model, whatever happens later in the procedure. Next, we estimate the $m - 1$ models with two additional regressors, one of which is the one already selected. Proceeding in this way, we always select a model in a well-defined and unambiguous manner. Notice however that in the final model there is no guarantee that all t-statistics are significant.

Second, the t-statistics are computed in the traditional manner, that is, using an estimate of σ^2 based on the submodel under consideration. In this way, we mimic precisely what happens in applied work. The critical value, however, is always taken to be 1.96. This does not make any serious difference, and is more in line with the normality assumptions made in the approximations.

We now discuss the effect of pretesting on the forecasts. The forecasts discussed below are one-period-ahead forecasts for the period 1992–2001, based on all information available at the moment of forecasting. For example, the forecast for the year 2000 is based on the model selected and estimated using the 1956–1999 data. It is thus possible (and indeed it happens) that the forecast in one year is based on a different model than in another year.

In Figure 1, the solid line gives the one-period-ahead forecasts $\hat{y}_{n+1}$, while the small open circles give the realized values y_{n+1}, the observed excess returns. The forecasts are the same, whether we take pretesting into account or not. The difference lies in the *distribution* of the forecasts. As a benchmark, the two dotted lines give the standard least-squares 95% prediction bounds (ignoring the effects of pretesting) as given in (5). These are the prediction bounds, reported when—as is usually the case—pretesting is ignored. They are symmetric around $\hat{y}_{n+1}$. We see that only 60% of the forecasts (six out of ten) lie in this standard prediction interval, which suggests that the bounds are too tight. The graphs for the period 1986–1991 are based on the in-sample estimates obtained for the full estimation period 1956–1991. The graphs for the period 1992–2001 are out-of-sample forecasts. The dash-dotted line gives the estimated bias-corrected forecast $\hat{y}_{n+1} + \psi_1(\hat{\boldsymbol{\theta}})$, based on (6) and Theorem 1, while the dashed lines show the conservative 95% confidence bounds of the bias-corrected forecast, given in (9), taking the inconsistency of $\hat{\boldsymbol{\theta}}$ into account.[3] In doing the calcula-

[3] To obtain ψ_1 (and ψ_2) we need to calculate an m-dimensional integral. For $m = 1$ and $m = 2$ we do this exactly (by quadrature); for $m \geq 3$ by Monte Carlo, based on 1000 replications. In each replication we generate a random sample according to the regression model (1) with $\boldsymbol{X}$ and $\boldsymbol{Z}$ the observed regressors, and perform forward model selection. The simulated results were tested for numerical stability.

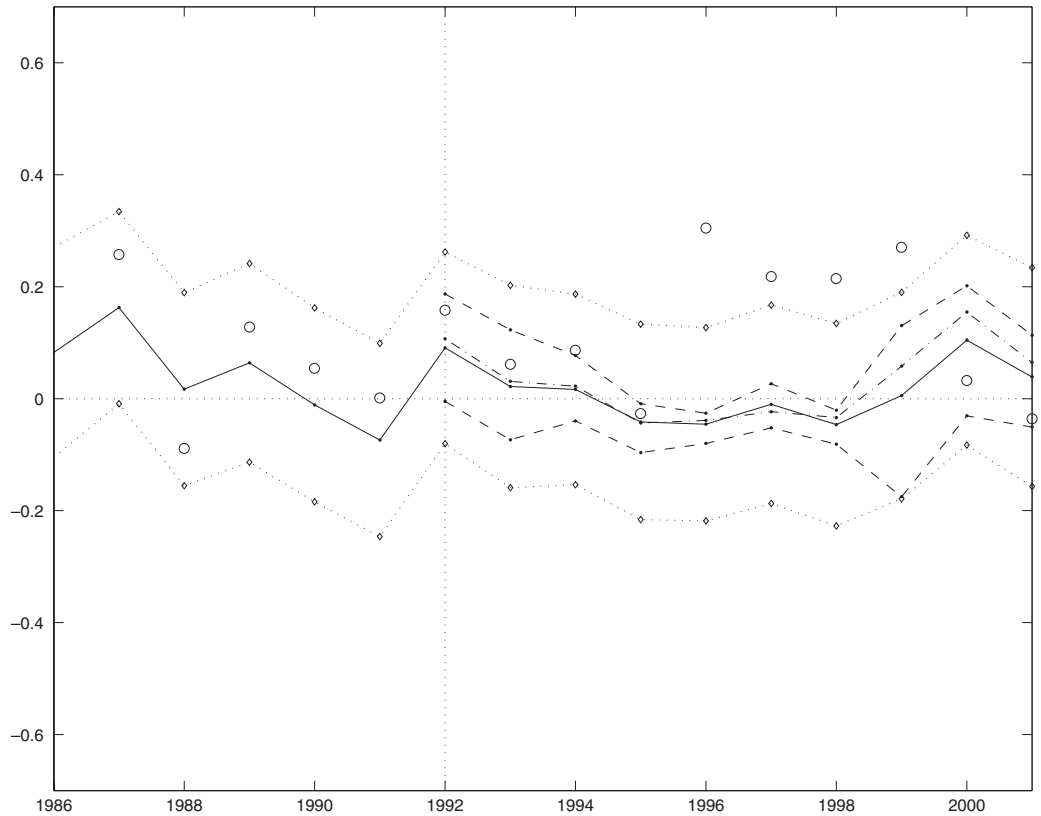

Figure 1. Pretest forecasts $\hat{y}_{n+1}$, bias-corrected forecasts and confidence bounds

tions for the dash-dotted and the dashed lines, we estimate σ^2 by the LS estimator in the unrestricted model, that is,

$$s_u^2 = \frac{1}{n-k-m}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}_u - \boldsymbol{Z}\hat{\boldsymbol{\gamma}})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}_u - \boldsymbol{Z}\hat{\boldsymbol{\gamma}}) \tag{12}$$

which simplifies the calculations without affecting the results; see later.

At first glance the estimated bias is small, because the dash-dotted line is quite close to the solid line. However, we have to take into account the inconsistency of $\hat{\boldsymbol{\theta}}$. The more conservative dashed bounds show that the bias can be large, at least in some years. In seven of the ten years the sign of the bias-corrected forecast is uncertain, thus showing that taking into account the pretest bias may matter a great deal.

In Figure 2, the solid line gives again the one-period-ahead forecasts, the small open circles are the realized values y_{n+1}, and the two dotted lines give the standard least-squares 95% prediction bounds ignoring the effects of pretesting. The two dash-dotted lines now show the approximate 95% prediction bounds of the pretest forecast, based on (10), while the dashed lines give the more conservative interval, based on (11). These intervals take account of both the bias and the variance.

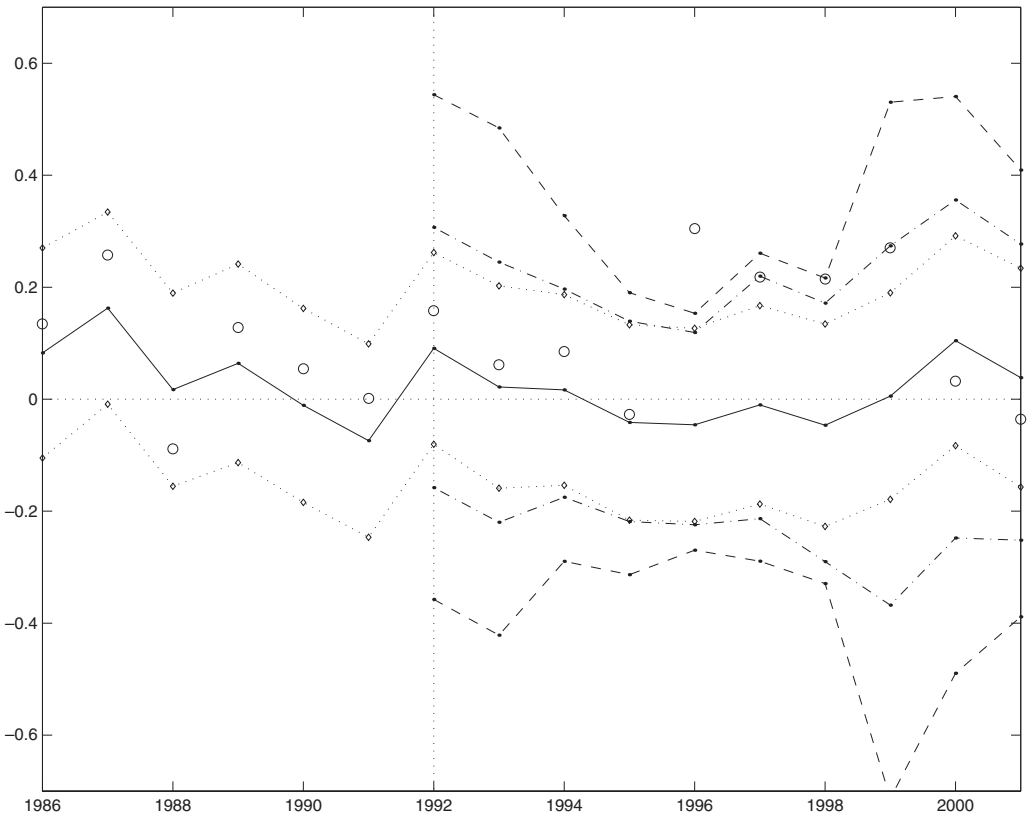

Figure 2. Pretest forecasts $\hat{y}_{n+1}$ with three sets of prediction intervals

Because of the bias effect, the intervals are not symmetric around $\hat{y}_{n+1}$. Now 80% of the forecasts lie in the approximate 95% prediction interval, and 90% in the more conservative interval. The year 1996 appears to be the most difficult to predict, partly because the market changed direction between 1995 and 1996. Although the pretest forecast is seriously biased in some years, and the standard deviation is seriously underestimated, and therefore standard prediction intervals can be very misleading for evaluating the accuracy of the forecast, it appears that the difference between the dotted and the dash-dotted lines is not spectacularly large, on average only 1.3 times as wide. Hence, ignoring the effects of pretesting on the prediction bounds of the forecast is not necessarily disastrous, at least in our application.

Lack of sensitivity in one direction does not, however, imply lack of sensitivity in another direction. One can argue that it is not the excess return but rather the *sign* of the excess return which is most relevant. Hence, as proposed by Granger and Pesaran (2000), we now study the forecast probability $\Pr(y_{n+1} > 0)$. Here the effect of ignoring pretesting will turn out to be rather more dramatic.

Since the error term is assumed to be normally distributed, we have

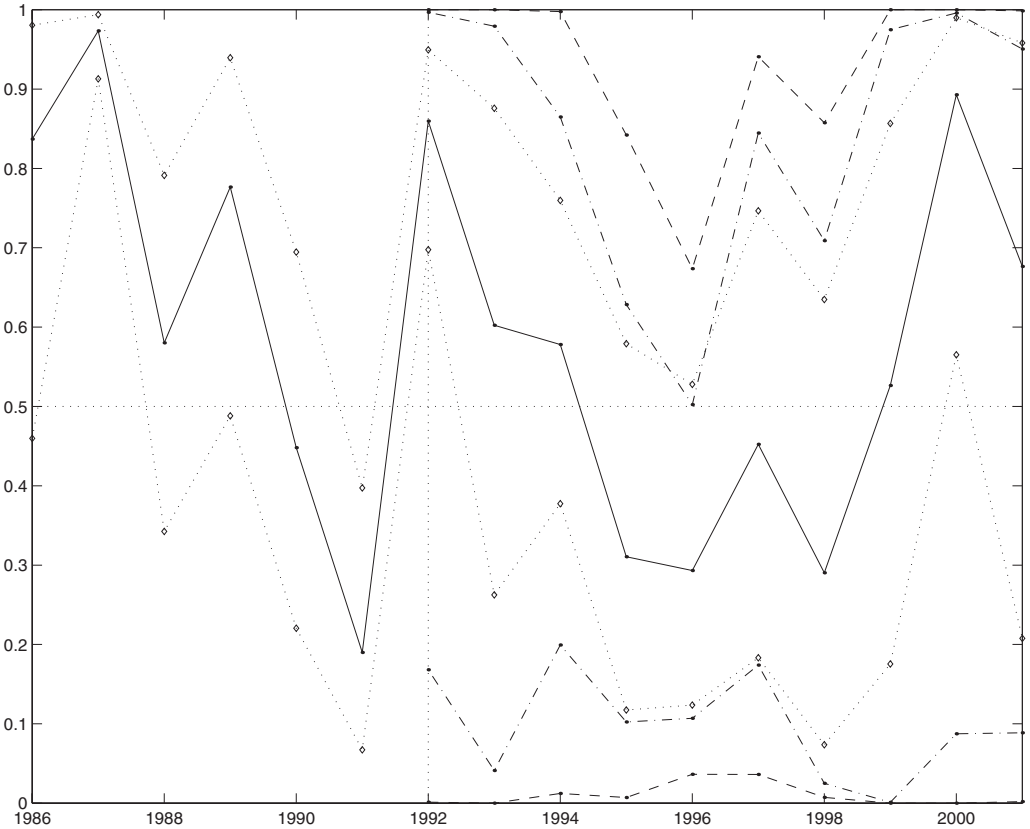

Figure 3. Pretest forecast probabilities $\Pr(y_{n+1} > 0)$ with three sets of prediction intervals

$$\begin{aligned}\Pr(y_{n+1} > 0) &= \Pr(\boldsymbol{x}'_{n+1}\boldsymbol{\beta} + \boldsymbol{z}'_{n+1}\boldsymbol{\gamma} + \varepsilon_{n+1} > 0) \\ &= \Pr(-\varepsilon_{n+1} < \boldsymbol{x}'_{n+1}\boldsymbol{\beta} + \boldsymbol{z}'_{n+1}\boldsymbol{\gamma}) \\ &= \Phi\left(\frac{\boldsymbol{x}'_{n+1}\boldsymbol{\beta} + \boldsymbol{z}'_{n+1}\boldsymbol{\gamma}}{\sigma}\right) \end{aligned} \tag{13}$$

where $\Phi(\cdot)$ denotes the standard normal c.d.f. If the value of $\Pr(y_{n+1} > 0)$ is larger than 0.5, the investor will conclude that the excess return is likely to be positive in the next period, and therefore will invest in stocks, if risk neutrality is assumed. If, on the other hand, the value of $\Pr(y_{n+1} > 0)$ is smaller than 0.5, the investor will conclude that excess return will be negative and will invest in bonds. Of course, the probability $\Pr(y_{n+1} > 0)$ is not known and needs to be estimated.

The solid line in Figure 3 gives the estimated probability $\hat{\Pr}(y_{n+1} > 0) = \Phi(\hat{y}_{n+1}/s_u)$. Since

$$\hat{y}_{n+1} > 0 \Leftrightarrow \Phi\left(\frac{\hat{y}_{n+1}}{s_u}\right) > \frac{1}{2} \Leftrightarrow \hat{\Pr}(y_{n+1} > 0) > \frac{1}{2}$$

we see that $\hat{\Pr}(y_{n+1} > 0)$ exceeds 0.5 if and only if $\hat{y}_{n+1}$ is positive, a fact confirmed by comparing the solid line in Figure 3 with the solid line in either Figure 1 or 2. If we take no account of the effects of pretesting, then a 95% prediction interval for the parameter $(\boldsymbol{x}'_{n+1}\boldsymbol{\beta} + \boldsymbol{z}'_{n+1}\boldsymbol{\gamma})/\sigma$ is given by

$$\frac{\hat{y}_{n+1}}{\sigma} \pm 1.96D, \qquad D := \sqrt{\boldsymbol{x}'_{n+1}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_{n+1} + \boldsymbol{\zeta}'_{n+1}\boldsymbol{W}\boldsymbol{\zeta}_{n+1}}$$

and hence

$$\Phi\left(\frac{\hat{y}_{n+1}}{s_u} - 1.96D\right) < \Pr(y_{n+1} > 0) < \Phi\left(\frac{\hat{y}_{n+1}}{s_u} + 1.96D\right)$$

denotes the estimated prediction interval for $\Pr(y_{n+1} > 0)$ when pretesting is not taken into account: the dotted lines in Figure 3. If, however, we do take account of pretesting, then

$$\hat{y}_{n+1} \sim (\boldsymbol{x}'_{n+1}\boldsymbol{\beta} + \boldsymbol{z}'_{n+1}\boldsymbol{\gamma} - \psi_1(\boldsymbol{\theta}), \psi_2(\boldsymbol{\theta})) \tag{14}$$

where ψ_1 and ψ_2 are defined in (6) and (7), so that an approximate 95% prediction interval for $\Pr(y_{n+1} > 0)$ is given by

$$\Phi\left(\frac{\hat{y}_{n+1} + \psi_1(\boldsymbol{\theta}) \pm 1.96\sqrt{\psi_2(\boldsymbol{\theta})}}{\sigma}\right) \tag{15}$$

This interval depends on $\boldsymbol{\theta}$ (and σ) which is unknown. We obtain an estimated prediction interval by replacing $\boldsymbol{\theta}$ and σ by the estimates $\hat{\boldsymbol{\theta}}$ and s_u, leading to the dash-dotted lines.

Finally, as in the previous section, we obtain more conservative bounds (the dashed lines), taking into account that $\hat{\boldsymbol{\theta}}$, while unbiased, is inconsistent:

$$\Phi\left(\frac{\hat{y}_{n+1} + \underline{C}_3(\hat{\boldsymbol{\theta}}, s_u)}{s_u}\right) < \Pr(y_{n+1} > 0) < \Phi\left(\frac{\hat{y}_{n+1} + \overline{C}_3(\hat{\boldsymbol{\theta}}, s_u)}{s_u}\right) \tag{16}$$

where

$$\underline{C}_3(\hat{\boldsymbol{\theta}}, \sigma) := \min_{\boldsymbol{\theta} \in \mathcal{H}(\hat{\boldsymbol{\theta}})} (\psi_1(\boldsymbol{\theta}) - 1.96\sqrt{\psi_2(\boldsymbol{\theta})})$$

and

$$\overline{C}_3(\hat{\boldsymbol{\theta}}, \sigma) := \max_{\boldsymbol{\theta} \in \mathcal{H}(\hat{\boldsymbol{\theta}})} (\psi_1(\boldsymbol{\theta}) + 1.96\sqrt{\psi_2(\boldsymbol{\theta})})$$

While the standard regression prediction intervals are already large, allowing only two years (1992, 2000) where a direction can be forecasted with any confidence, the pretest prediction intervals are such that we cannot be confident in *any* year. This is true for the dash-dotted lines and, *a forteriori*, for the more conservative dashed lines. The difference between the dotted and the dash-dotted lines is twice as large as in Figure 2, on average 2.6 times as wide.

We conclude that ignoring the effects of pretesting on the distribution of the forecast can lead to a serious misrepresentation. The pretest forecast is biased and has a larger variance than is apparent from the regression results. The one-period-ahead forecasts are much less precise than naive econometrics would lead us to believe. The effects of pretesting on forecasting are thus potentially serious and should be analysed and incorporated in econometric analyses.

'OPTIMAL' FORECASTS USING LAPLACE WEIGHTS

We have seen that in evaluating the properties of forecasts, especially forecast probabilities, we need to take the model selection aspect into account. So far, we have only considered the standard pretest procedure, where we first select the 'best' model and then forecast on the basis of this selected model. Such a procedure is discontinuous and hence inadmissible. Since we are not in the business of finding the 'best' model, but rather of finding the 'best' forecast, we may wish to consider a (continuous) weighted average of models instead of the (discontinuous) pretest model selection, somewhat in the spirit of 'thick modelling' (see Granger and Jeon, 2001). But which weights should be taken? What constitutes an optimal forecast depends on the forecast user. The Laplace weights, introduced by Magnus (2002), are 'optimal' in the sense that they lead to estimates and forecasts with a low MSE (MSFE). Low MS(F)E seems a reasonable criterion, although one might argue in our example that an optimal forecast is one where wealth is maximized (as is the main concern in PT94), or the number of correct sign predictions, or model stability (as argued by Paye and Timmermann, 2002). Let us compare the standard pretest (forward) procedure with the restricted, unrestricted and Laplace procedures, keeping these different notions of optimality in mind.

The Laplace weights were introduced by Magnus (2002) for the case $m = 1$ in the estimation context. When $m = 1$, there are only two possible models, the restricted (r) and the unrestricted (u), and the forecast takes the simple form (see (3))

$$\hat{y}_{n+1} = \lambda \hat{y}_{n+1}^{(u)} + (1-\lambda)\hat{y}_{n+1}^{(r)}$$

The proposed weight function $\lambda = \lambda(\hat{\theta})$ is

$$\lambda(\hat{\theta}) = \frac{\int \theta \pi(\theta) \exp\left(-(\hat{\theta}-\theta)^2 \big/ (2\sigma^2)\right) d\theta}{\hat{\theta} \int \pi(\theta) \exp\left(-(\hat{\theta}-\theta)^2 \big/ (2\sigma^2)\right) d\theta}$$

where the prior π is the 'neutral' Laplace density

$$\pi(\theta) = \frac{c}{2}\exp(-c|\theta|), \qquad -\infty < \theta < \infty, \qquad c = \sigma \log 2$$

The neutrality of the prior guarantees that median $(\theta) = 0$ and median$(\theta^2) = \sigma^2$. When $m > 1$ it is not so clear how the weights should be taken. However, in the special case where $\boldsymbol{Z'MZ} = \boldsymbol{I}_m$, the multi-dimensional problem separates into m one-dimensional problems, and we can use the Laplace weights for each dimension separately; see DM04, theorem 2.

Let us consider the 'orthogonalization' $\boldsymbol{Z'MZ} = \boldsymbol{I}_m$ in some more detail. Orthogonalization can always be achieved by taking appropriate linear combinations of the m auxiliary regressors in $\boldsymbol{Z}$

(leaving the focus regressors unchanged). More specifically, let $\boldsymbol{T}_1$ be an orthogonal $m \times m$ matrix such that $\boldsymbol{T}_1'\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z}\boldsymbol{T}_1 = \boldsymbol{\Lambda}$ (diagonal). Then, letting $\boldsymbol{T} = \boldsymbol{T}_1\boldsymbol{\Lambda}^{-1/2}$, we have $\boldsymbol{T}'\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z}\boldsymbol{T} = \boldsymbol{I}_m$. Now define new auxiliary regressors $\boldsymbol{Z}^* = \boldsymbol{Z}\boldsymbol{T}$ and $\boldsymbol{z}_{n+1}^* = \boldsymbol{T}'\boldsymbol{z}_{n+1}$. Then, clearly, $\boldsymbol{Z}^{*\prime}\boldsymbol{M}\boldsymbol{Z}^* = \boldsymbol{I}_m$. As a consequence of this transformation, $\boldsymbol{\zeta}_{n+1}$, $\boldsymbol{R}(\boldsymbol{\theta}) := \mathrm{MSE}(\boldsymbol{W}\hat{\boldsymbol{\theta}})$ and MSFE will all change, but $\boldsymbol{\zeta}_{n+1}'\boldsymbol{\zeta}_{n+1}$ will not change. This follows because

$$\begin{aligned}\boldsymbol{\zeta}_{n+1} &= \boldsymbol{Q}'\boldsymbol{x}_{n+1} - (\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z})^{-1/2}\boldsymbol{z}_{n+1} \\ &= (\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z})^{-1/2}\left(\boldsymbol{Z}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_{n+1} - \boldsymbol{z}_{n+1}\right)\end{aligned}$$

so that

$$\begin{aligned}\boldsymbol{\zeta}_{n+1}^* &= (\boldsymbol{Z}^{*\prime}\boldsymbol{M}\boldsymbol{Z}^*)^{-1/2}\left(\boldsymbol{Z}^{*\prime}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_{n+1} - \boldsymbol{z}_{n+1}^*\right) \\ &= \boldsymbol{T}'\left(\boldsymbol{Z}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_{n+1} - \boldsymbol{z}_{n+1}\right)\end{aligned}$$

Then the fact that $\boldsymbol{T}\boldsymbol{T}' = (\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z})^{-1}$ implies that $\boldsymbol{\zeta}_{n+1}^{*\prime}\boldsymbol{\zeta}_{n+1}^* = \boldsymbol{\zeta}_{n+1}'\boldsymbol{\zeta}_{n+1}$. The only difference between

$$\begin{aligned}\mathrm{MSFE} &= \sigma^2\left(\boldsymbol{x}_{n+1}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_{n+1} + 1\right) + \boldsymbol{\zeta}_{n+1}'\boldsymbol{R}(\boldsymbol{\theta})\boldsymbol{\zeta}_{n+1} \\ &= (\boldsymbol{\zeta}_{n+1}'\boldsymbol{\zeta}_{n+1})\left(\frac{\sigma^2\left(\boldsymbol{x}_{n+1}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_{n+1} + 1\right)}{\boldsymbol{\zeta}_{n+1}'\boldsymbol{\zeta}_{n+1}} + \frac{\boldsymbol{\zeta}_{n+1}'\boldsymbol{R}(\boldsymbol{\theta})\boldsymbol{\zeta}_{n+1}}{\boldsymbol{\zeta}_{n+1}'\boldsymbol{\zeta}_{n+1}}\right)\end{aligned}$$

and

$$\mathrm{MSFE}^* = \left(\boldsymbol{\zeta}_{n+1}^{*\prime}\boldsymbol{\zeta}_{n+1}^*\right)\left(\frac{\sigma^2\left(\boldsymbol{x}_{n+1}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_{n+1} + 1\right)}{\boldsymbol{\zeta}_{n+1}^{*\prime}\boldsymbol{\zeta}_{n+1}^*} + \frac{\boldsymbol{\zeta}_{n+1}^{*\prime}\boldsymbol{R}^*(\boldsymbol{\theta})\boldsymbol{\zeta}_{n+1}^*}{\boldsymbol{\zeta}_{n+1}^{*\prime}\boldsymbol{\zeta}_{n+1}^*}\right)$$

lies in the two expressions

$$\xi^2 := \frac{\boldsymbol{\zeta}_{n+1}'\boldsymbol{R}(\boldsymbol{\theta})\boldsymbol{\zeta}_{n+1}}{\boldsymbol{\zeta}_{n+1}'\boldsymbol{\zeta}_{n+1}} \quad \text{and} \quad \xi^{*2} := \frac{\boldsymbol{\zeta}_{n+1}^{*\prime}\boldsymbol{R}^*(\boldsymbol{\theta})\boldsymbol{\zeta}_{n+1}^*}{\boldsymbol{\zeta}_{n+1}^{*\prime}\boldsymbol{\zeta}_{n+1}^*}$$

At first sight, the difference between ξ^2 and ξ^{*2}, and hence between MSFE and MSFE*, may seem trivial. This, however, is not so. First, while MSFE depends on the model selection procedure [for example, forward (specific-to-general) or backward (general-to-specific)], MSFE* is independent of the selection procedure. Second, while the eigenvalues of $\boldsymbol{R}(\boldsymbol{\theta})$ are not necessarily bounded, the eigenvalues of $\boldsymbol{R}^*(\boldsymbol{\theta})$ are always bounded, so that ξ^{*2} is always finite even when ξ^2 is infinite.[4] Third, simple analytical expressions exist for the MSFE*, but not for MSFE. And finally, the 'optimal' WALS forecast can be applied quite easily in the case of MSFE*, but not in the case of MSFE.

We now compare the 'forward', 'unrestricted' and 'Laplace' procedures. The forward pretest procedure was already discussed and applied; it is the standard procedure used in applied work. The

[4] In the forward pretest procedure, ξ^2 can become as large as we please by making $\boldsymbol{M}\boldsymbol{z}_i$ and $\boldsymbol{M}\boldsymbol{z}_j$ more and more correlated; see DM04, section 6.

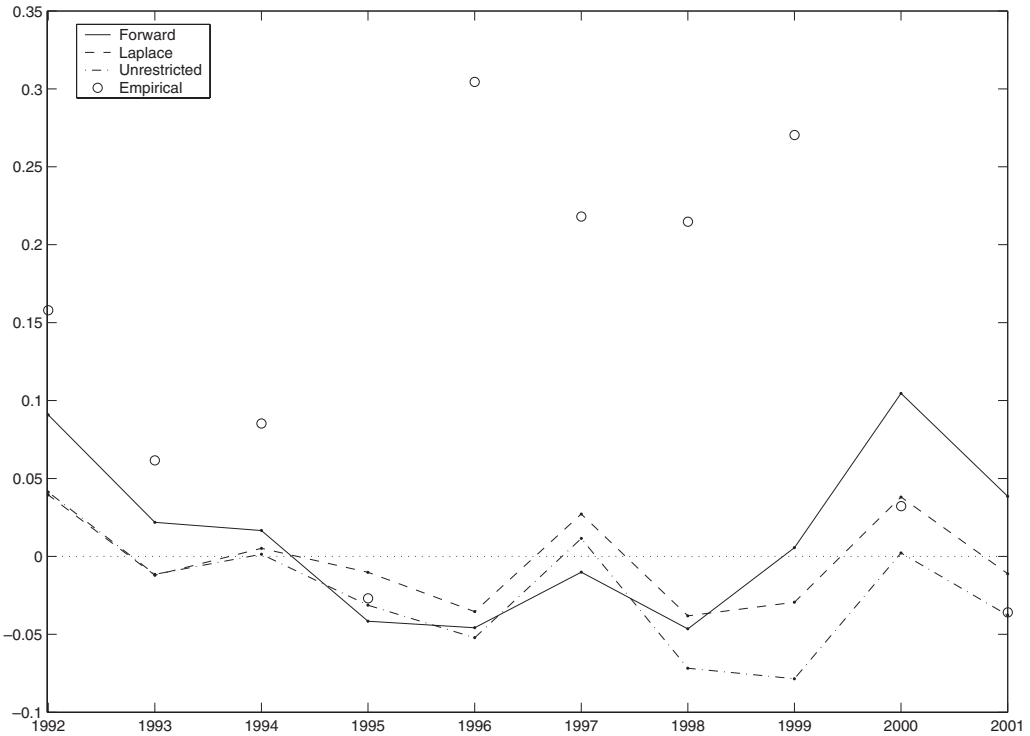

Figure 4. Point forecasts for three model selection procedures

unrestricted (true) model is the one where all auxiliary regressors are included. The Laplace procedure first transforms the auxiliary regressors $\boldsymbol{Z}$ so that they become 'orthogonal' (in the sense that $\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z} = \boldsymbol{I}_m$). The Laplace weights λ_i determine how much weight is attached to each auxiliary regressor, essentially depending on the relevant t-statistic. The Laplace procedure can thus be viewed as a continuous version of the discrete (and hence inadmissible) pretest procedure.

In Figure 4 we plot the point forecasts for the three procedures. The solid line and the small open circles in Figure 4 correspond to the forward procedure and are the same as in Figures 1 and 2. But now we plot the forecasts for two other procedures, unrestricted and Laplace, as well. None of the three procedures predict particularly well. The Laplace forecasts and the unrestricted forecasts move closely together. In particular, their sign is always the same in our application. For each of the three procedures six out of ten predictions have the correct sign. Thus, the point forecasts themselves do not give much indication of which procedure is best.

In Figure 5 we plot the forecast probabilities for the three procedures. The solid line is the same as in Figure 3. The triangles depict the direction of the market: down in 1995 and 2001, up in the other eight years. For example, in 1992, all three procedures predicted correctly that the market would go up. In 1996 and 1998 the market went up, but all three procedures predicted that it would go down. In 2001, the Laplace and unrestricted procedures correctly predicted the crash, but the forward procedure did not.

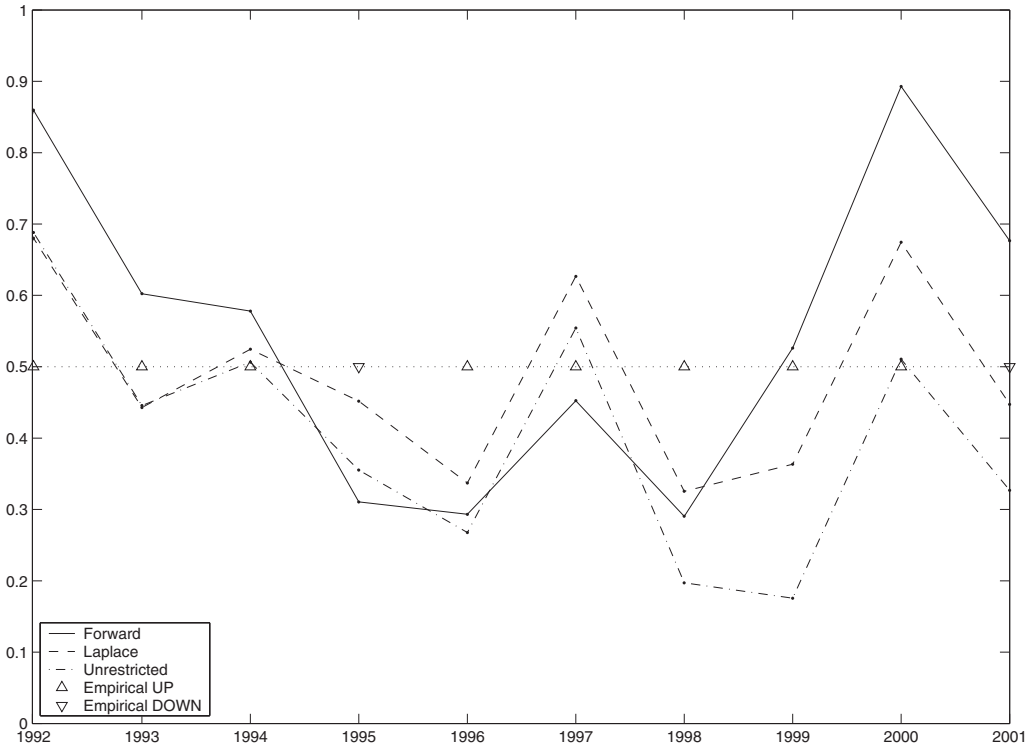

Figure 5. Forecast probabilities for three procedures

Further insight into the relative performance of the three procedures is obtained from Table I. A '+' indicates that the market is predicted to go up and a '–' that it goes down. The best procedure is the one where each prediction is correct. Each procedure predicts 60% correctly. The Laplace procedure is best in the sense that it has the lowest MSFE. The forward procedure has the highest mean return (expected profit), but also the highest standard deviation (volatility). The Sharpe index—combining the mean and the standard deviation of the return—is highest for the Laplace and unrestricted procedures. Hence, they provide higher expected profit for the same level of volatility. We conclude that the Laplace procedure performs well, both in terms of MSFE and in terms of Sharpe index. However, the amount of improvement is relatively small.

As a final comparison between the three procedures, let us consider the MSFE of the WALS estimator, given in Theorem 1:

$$\begin{aligned}\mathrm{MSFE} &= \sigma^2\left(\boldsymbol{x}_{n+1}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_{n+1}+1\right)+\boldsymbol{\zeta}_{n+1}'\mathrm{MSE}(\boldsymbol{W}\hat{\boldsymbol{\theta}})\boldsymbol{\zeta}_{n+1}\\ &= \psi_2(\boldsymbol{\theta})+\psi_1^2(\boldsymbol{\theta})+\sigma^2\end{aligned}$$

This expression depends on $\boldsymbol{\theta}$ (and σ^2), which is unknown. Following the same approach as before, we obtain a 95% bound for the MSFE as

Table I. Performance of three model selection procedures

Year	Model selection procedure			Best
	Forward	Laplace	Unrestricted	
1992	+	+	+	+
1993	+	−	−	+
1994	+	+	+	+
1995	−	−	−	−
1996	−	−	−	+
1997	−	+	+	+
1998	−	−	−	+
1999	+	−	−	+
2000	+	+	+	+
2001	+	−	−	−
MSFE (%)	3.35	3.33	4.01	
Mean return (%)	10.64	9.87	9.87	18.38
S.d. of return (%)	9.49	8.12	8.12	12.01
Sharpe's index	0.603	0.609	0.609	1.121

$$\text{MSFE} < C_4(\hat{\boldsymbol{\theta}}, \hat{\sigma}) + \hat{\sigma}^2$$

where

$$C_4(\hat{\boldsymbol{\theta}}, \sigma) := \max_{\boldsymbol{\theta} \in \mathcal{H}(\hat{\boldsymbol{\theta}})} (\psi_2(\boldsymbol{\theta}) + \psi_1^2(\boldsymbol{\theta}))$$

In Figure 6 we compare the bounds of the MSFE for the forward, pretest and Laplace procedures.

Figure 6 shows rather more convincingly the dangers of the forward procedure and the superiority of the Laplace procedure. The MSFE bound of the Laplace forecast is very much lower than for the pretest forecast, and uniformly so. Moreover, if we compare the bounds with their theoretical minimum σ^2 ($\boldsymbol{x}'_{n+1}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_{n+1} + 1$), then the difference between the procedures becomes even more pronounced. We also observe that the MSFE bounds vary significantly over time.

Based on these results we conclude, be it tentatively, that—if our focus is forecasting rather than model selection—better forecasts can be generated using the Laplace weights than using the traditional pretest forecasts.

CONCLUDING REMARKS

On the basis of our theoretical and empirical results, we conclude that taking explicit account of pretesting in assessing the properties of one-period-ahead forecasts is essential in econometrics, if we wish to be credible to policy makers and others.

We all know that we use the same data for model selection and forecasting (and estimation), that therefore pretesting takes place, and hence that the properties of forecasts (and estimators) are affected. This paper shows that it is possible to take pretesting into proper account, and that it matters.

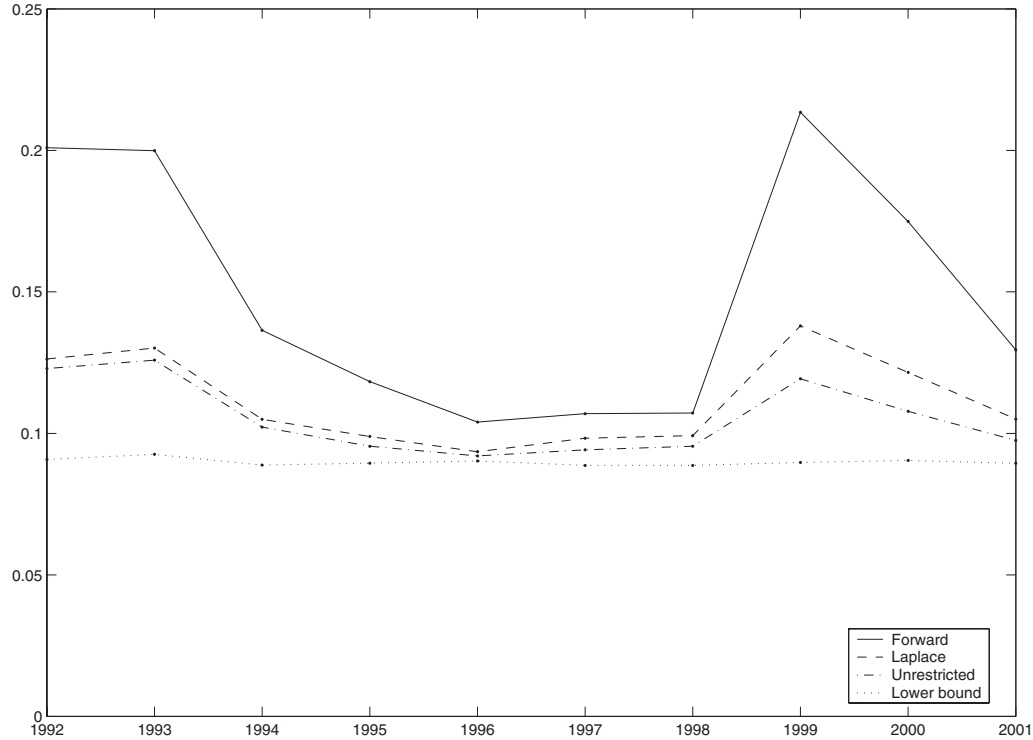

Figure 6. Upper bounds of MSFE, σ^2 estimated by (12)

In addition, we show that an alternative exists to the (discontinuous, hence inadmissible) traditional pretest procedure, based on Laplace weights. These weights have good theoretical properties, and they appear to behave well in practice too.

Finally, let us consider the effect of estimating σ^2. So far we derived the prediction intervals on the assumption that σ^2 is known, and only at the final stage did we substitute σ^2 by its estimate (12), based on the unrestricted model.

We now want to treat σ^2 'properly', that is, we estimate it by the LS estimate of the selected submodel

$$s^2_{(i)} = \frac{1}{n-k-m+r_i}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}_{(i)} - \boldsymbol{Z}\boldsymbol{c}_{(i)})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}_{(i)} - \boldsymbol{Z}\boldsymbol{c}_{(i)})$$

and we take its distribution into account when selecting the model. There is no theoretical problem in doing the calculations, because the estimator for σ^2 will depend on $\boldsymbol{M}\boldsymbol{y}$, so that Theorem 1 still applies, but they are much more complicated and time-consuming.

In Figure 7 we recalculate the MSFE bounds of Figure 6, but now taking the estimation of σ^2 into proper account. As the plots show, the difference between Figures 6 and 7 is very small. This confirms the conclusion in Danilov (2002) that all qualitative (and most quantitative) results are not affected when we ignore the obvious fact that σ^2 is not known.

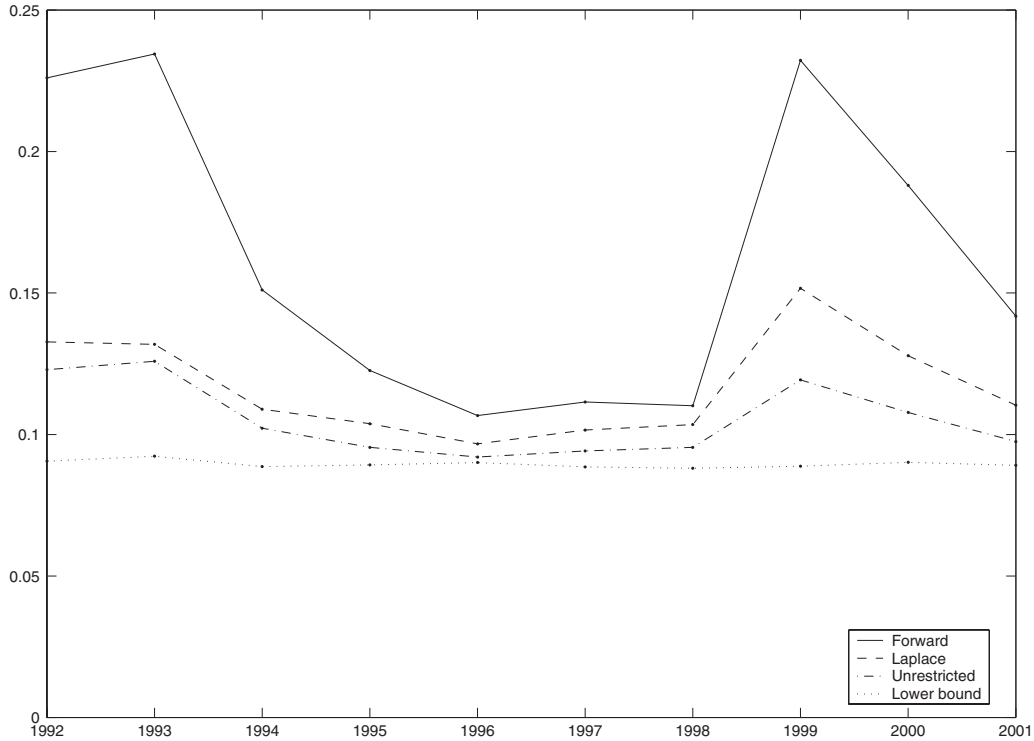

Figure 7. Upper bounds of MSFE, σ^2 estimated 'properly'

DATA APPENDIX

We attempted to use the same data as in PT94 (Pesaran and Timmermann, 1994), but could not quite do so for four reasons. First, the data set used by PT94 is not available now. We had access to the data used by PT95 (Pesaran and Timmermann, 1995); not, however, to the original data, but the data recently updated by the authors. We could not obtain all the data for 1954 and 1955, because SPREAD and YSP are not available in 1953 and 1954. Second, our data set extends to the year 2001, so that we had to employ a slightly different regressor instead of the term premium, namely SPREAD instead of TERM, since the 6-month commercial paper rate is no longer published by the Federal Reserve. Third, we had no access to the CRSP (Center for Research in Security Prices) tapes, in particular not to the Fama–Bliss risk-free rates files, that were used by Pesaran and Timmermann. Therefore an alternative source had to be used. Finally, various business cycle indicators employed in PT94 are in fact composite indices, subject to revisions and renormalizations. The indices that agree with the Citybase definition (used in PT94) end in November 1995, and a slightly different definition was employed afterwards. In this appendix we describe briefly how the data are constructed. Tables AI and AII provide the full data set employed.

Table AIa. Dependent variable and focus regressors, 1956–1991

Year	y_t	PI_{t-2}	$\mathrm{DI3}_{t-1}$	SPREAD_{t-1}
1956	0.2281	0.0022	0.12	0.13
1957	0.0365	0.0022	0.17	0.43
1958	–0.0572	0.0275	0.26	0.26
1959	0.3504	0.0366	–2.05	1.68
1960	0.0054	0.0221	0.09	0.36
1961	0.0997	–0.0020	–0.01	0.91
1962	0.1191	0.0083	0.16	0.44
1963	–0.0403	–0.0002	0.45	0.32
1964	0.1679	0.0032	0.21	0.25
1965	0.1311	–0.0035	0.02	0.33
1966	0.0511	0.0037	0.32	0.14
1967	–0.0858	0.0175	0.57	0.20
1968	0.0526	0.0308	–0.71	0.99
1969	0.0953	0.0122	0.32	0.59
1970	–0.2049	0.0286	0.71	–0.05
1971	0.0847	0.0359	0.87	0.28
1972	0.0730	0.0341	–1.69	0.93
1973	0.1054	0.0297	–0.96	0.61
1974	–0.2000	0.0306	0.92	–0.12
1975	–0.2344	0.0876	0.12	0.50
1976	0.2972	0.1425	–2.12	1.65
1977	–0.0034	0.1024	–0.78	0.43
1978	–0.1354	0.0430	–0.17	–0.08
1979	0.1058	0.0630	0.24	0.27
1980	0.1011	0.0753	0.54	0.81
1981	0.0726	0.1054	–0.12	–0.04
1982	–0.1544	0.1263	1.88	–0.10
1983	0.1283	0.0885	–0.23	0.04
1984	0.0863	0.0393	0.20	–0.14
1985	0.0500	0.0160	0.38	0.19
1986	0.1346	0.0207	–0.96	–0.24
1987	0.2575	0.0086	–0.22	0.55
1988	–0.0890	–0.0134	0.40	–0.01
1989	0.1278	0.0203	0.37	0.98
1990	0.0546	0.0247	1.03	0.39
1991	0.0011	0.0500	–0.03	0.16

Dependent variable

The dependent variable y_t denotes the excess return in year t, and is defined by

$$y_t = \mathrm{NRSP}_t - \mathrm{I12}_{t-1}$$

where

$$\mathrm{NRSP}_t = \frac{\mathrm{PSP}_t - \mathrm{PSP}_{t-1} + \mathrm{DIVSP}_{t-1}}{\mathrm{PSP}_{t-1}}$$

Table AIb. Dependent variable and focus regressors, 1992–2001

Year	y_t	PI_{t-2}	$DI3_{t-1}$	$SPREAD_{t-1}$
1992	0.1580	0.0481	–0.92	0.73
1993	0.0616	0.0214	–0.98	0.15
1994	0.0852	0.0122	–0.06	0.42
1995	–0.0269	0.0124	–0.07	0.15
1996	0.3045	0.0063	0.80	0.27
1997	0.2181	0.0190	–0.41	0.34
1998	0.2147	0.0258	–0.01	0.30
1999	0.2703	0.0041	–0.01	0.38
2000	0.0323	–0.0088	0.14	0.44
2001	–0.0360	0.0180	0.56	0.28

denotes the annual rate of return on the SP 500 index, and $I12_{t-1}$ denotes the 12-month T-bill rate on the last trading day of January in the year $t - 1$.

The variable I12 is obtained from PT95, up to the year 1992. Later years are obtained from the H15 Federal Reserve Statistical Release, section Weekly Releases, Selected Interest Rates, Historical data, Treasury bills, Secondary market, 1-year, Business.[5]

The variable PSP denotes the nominal price index for the SP 500 portfolio at the close of the last trading day of January. *Sources*: PT95 (for the years 1955–1992) and DataStream (from 31 December 1964 up to 2001). We used the PT95 data set updated from DataStream where necessary.

DIVSP denotes the average nominal dividends per share for the SP 500 portfolio paid during the calendar year. It is constructed as DIVSP = PSP × YSP, where YSP is defined below.

Focus regressors

The first focus regressor is the constant term. In addition, we have three other focus regressors. The second is PI, the annual inflation rate, computed as $PI_t = \log(PPIAV_t/PPIAV_{t-1})$, where PPIAV denotes the annual average of the producer price index (PPI, finished goods). *Source*: website of the US Department of Labor, Bureau of Labor Statistics, Series: Producer Price Index by Finished Goods (April 1947 to present).[6]

The third is DI3, the change in the 3-month T-bill rate, defined as the difference between the 3-month T-bill rate in January (I3:JAN) and the 3-month T-bill rate in October (I3:OCT) of the previous year, both measured at the last trading day of the month. *Source*: H15 Federal Reserve Statistical Release, section Weekly Releases, Selected Interest Rates, Historical data, Treasury bills, Secondary market, 3-month, Business.[7]

The fourth focus regressor is SPREAD, the credit spread, defined as the difference between the 3-month financial paper rate (IF3:JAN) and I3:JAN. PT94 employ the term premium, that is the difference between the 6-month commercial paper rate and the 3-month T-bill rate in January. Since the 6-month commercial paper rate does not exist after 1997, we use the 3-month financial paper rate instead. The 3-month financial paper rate data consist of two files, before September 1997 and after. *Sources*: H15 Federal Reserve Statistical Release, section Weekly Releases, Selected Interest Rates, Historical data, Finance paper placed directly (historical), 3-month, Monthly (1955–1997),

[5] See http://www.federalreserve.gov/releases/h15/data/b/tbsm1y.txt
[6] Available online at http://www.bls.gov/
[7] See http://www.federalreserve.gov/Releases/h15/data/b/tbsm3m.txt

Table AIIa. Auxiliary regressors, 1956–1991

Year	YSP_{t-1}	DIP_{t-1}	PER_{t-1}	$DLEAD_{t-2}$
1956	0.0510	136.2792	–0.0583	0.9986
1957	0.0398	148.1861	0.1204	1.0655
1958	0.0382	157.7279	0.0423	0.9949
1959	0.0391	155.4372	0.0141	0.9678
1960	0.0387	190.3797	–0.0670	1.0000
1961	0.0311	208.2837	0.1129	1.0492
1962	0.0348	202.3910	0.0223	0.9823
1963	0.0300	258.8171	0.0066	1.0232
1964	0.0333	212.3313	0.0800	1.0189
1965	0.0318	216.9554	0.0595	1.0235
1966	0.0297	223.2941	0.0652	1.0266
1967	0.0298	215.8158	0.0947	1.0247
1968	0.0333	185.6376	0.0848	1.0034
1969	0.0320	206.3934	0.0214	0.9931
1970	0.0307	210.8557	0.0541	1.0219
1971	0.0318	186.6205	0.0454	1.0045
1972	0.0382	198.8993	–0.0335	0.9573
1973	0.0318	222.2841	0.0136	1.0329
1974	0.0285	222.3302	0.0925	1.0477
1975	0.0301	161.8866	0.0781	1.0108
1976	0.0428	106.8587	–0.0151	0.9431
1977	0.0435	142.4935	–0.0916	0.9659
1978	0.0376	130.9383	0.0876	1.0695
1979	0.0446	105.7509	0.0783	1.0165
1980	0.0516	101.5938	0.0570	1.0108
1981	0.0526	91.4693	0.0327	0.9871
1982	0.0510	101.5335	–0.0280	0.9620
1983	0.0499	96.4215	0.0162	1.0023
1984	0.0575	115.2959	–0.0552	0.9809
1985	0.0440	148.3745	0.0366	1.0746
1986	0.0457	124.5979	0.0857	1.0182
1987	0.0419	155.8990	0.0163	1.0000
1988	0.0345	209.0400	0.0112	1.0252
1989	0.0303	204.4862	0.0453	1.0235
1990	0.0349	150.9822	0.0444	1.0010
1991	0.0326	164.9125	0.0178	0.9960

and H15 Federal Reserve Statistical Release, section Weekly Releases, Selected Interest Rates, Historical data, Commercial paper (Financial), 3-month, Monthly (1997–2002).[8]

Auxiliary regressors

We consider four auxiliary regressors. First, YSP, the dividend yield on the SP 500 portfolio, is defined as $YSP_t = DIVSP_{t-1}/PSP_t$. *Sources*: PT95, datafile (1955–1992) and DataStream (from January 1965 to present).

Second, DIP, the annual change in industrial production, is computed as $DIP_t = \log(IPAV_t/IPAV_{t-1})$, where IPAV is the 12-month average of the industrial production index (IP). *Source*: On-

[8] See http://www.federalreserve.gov/Releases/h15/data/m/hfp3m.txt for the historical data and . . . /fp3m.txt for the recent data.

Table AIIb. Auxiliary regressors, 1992–2001

Year	YSP_{t-1}	DIP_{t-1}	PER_{t-1}	$DLEAD_{t-2}$
1992	0.0349	186.9279	–0.0021	0.9880
1993	0.0326	264.9310	–0.0202	0.9868
1994	0.0297	291.4800	0.0309	1.0103
1995	0.0277	278.0400	0.0343	1.0071
1996	0.0283	205.2000	0.0527	1.0293
1997	0.0255	211.4400	0.0469	0.9910
1998	0.0218	257.5200	0.0447	1.0117
1999	0.0176	288.6000	0.0650	1.0195
2000	0.0147	391.6800	0.0466	1.0169
2001	0.0125	304.9200	0.0406	1.0215

line database of the Federal Reserve Bank of St. Louis.[9] The data are monthly, seasonally adjusted, and range from January 1940 to August 2001. The data series is an index, base year 1992.

Third, PER, the price–earnings ratio for the SP 500 index, is the ratio of the price of stock to the earnings of companies per unit of stock. We have two sources for these variables, one from PT95, the other from DataStream. (Note that PT95 give the earnings–price ratio, rather than the price–earnings ratio.) DataStream use the annualized price–earnings ratio.

Finally, DLEAD denotes the annual change in the leading business cycle indicator, and is defined as $DLEAD_t = \log(LEAD_t/LEAD_{t-1})$. Here, LEAD is the 12-month average of a composite of 11 leading business cycle indicators. The leading indicator LEAD is taken from the data set BCIH-01.dat (composite indexes), distributed together with BCI Data Manager (January 1948 to November 1995).[10] For more recent data we extend the series as follows. We take the 'updated series' from the Economagic website.[11] This series is, however, calculated using a slightly different definition and base year. Therefore, we regress the old series on the updated series over the period where they overlap ($R^2 = 0.99$), and use the intercept and slope estimates and the values of the updated series to predict the missing years of the old series.

ACKNOWLEDGEMENTS

Preliminary versions of this paper were presented at Tilburg University, the University of Cambridge (UK) and at ESEM 2002 in Venice, 25–28 August 2002. We are grateful to seminar participants, to Hashem Pesaran, two referees and the editor for constructive and very useful comments.

REFERENCES

Abadir KM, Magnus JR. 2002. Notation in econometrics: a proposal for a standard. *The Econometrics Journal* **5**: 76–90.

Balvers RJ, Cosimano TF, McDonald B. 1990. Predicting stock returns in an efficient market. *Journal of Finance* **45**: 1109–1128.

[9] See http://www.stls.frb.org/

[10] See http://www.wfu.edu/~cottrell/bci/Software.html

[11] See http://www.economagic.com/em-cgi/data.exe/feddal/jlead

Black F, Jensen MC, Scholes M. 1972. The capital asset pricing model: some empirical tests. In *Studies in the Theory of Capital Markets*, Jensen MC (ed.). Praeger: New York; 79–121.
Chen NF, Roll R, Ross SA. 1986. Economic forces and the stock market. *Journal of Business* **59**: 382–403.
Cheng TCE, Lo YK, Ma KW. 1990. Forecasting stock price index by multiple regression. *Managerial Finance* **16**: 27–32.
Danilov D. 2002. Estimation of the mean of a univariate normal distribution when the variance is not known. CentER Discussion Paper, Tilburg University, The Netherlands, submitted for publication.
Danilov D, Magnus JR. 2004. On the harm that ignoring pretesting can cause. *Journal of Econometrics*, in press.
Fama EF, French KR. 1989. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* **25**: 23–49.
Fama EF, French KR. 1992. The cross-section of expected stock returns. *Journal of Finance* **47**: 427–465.
Fama EF, MacBeth JD. 1973. Risk, return and equilibrium: empirical tests. *Journal of Political Economy* **81**: 607–636.
Foster FD, Smith T, Whaley RE. 1997. Assessing goodness-of-fit of asset pricing models: the distribution of the maximal R^2. *Journal of Finance* **52**: 591–607.
French KR, Schwert GW, Stambaugh R. 1987. Expected stock returns and volatility. *Journal of Financial Economics* **19**: 3–30.
Granger CWJ, Jeon Y. 2001. Thick modeling. Department of Economics, University of California at San Diego, Working Paper.
Granger CWJ, Pesaran MH. 2000. Economic and statistical measures of forecast accuracy. *Journal of Forecasting* **19**: 537–560.
Lo AW, MacKinlay AC. 1990. Data-snooping biases in tests of financial asset pricing models. *Review of Financial Studies* **3**: 431–467.
Magnus JR. 2002. Estimation of the mean of a univariate normal distribution with known variance. *The Econometrics Journal* **5**: 225–236.
Magnus JR, Durbin J. 1999. Estimation of regression coefficients of interest when other regression coefficients are of no interest. *Econometrica* **67**: 639–643.
Paye BS, Timmermann A. 2002. How stable are financial prediction models? Evidence from US and international stock market data. Department of Economics, University of California at San Diego, Working Paper.
Pesaran MH, Timmermann A. 1994. Forecasting stock returns. An examination of stock market trading in the presence of transaction costs. *Journal of Forecasting* **13**: 335–367.
Pesaran MH, Timmermann A. 1995. Predictability of stock returns: robustness and economic significance. *Journal of Finance* **50**: 1201–1228.
Rozeff MS. 1984. Dividend yields are equity risk premiums. *Journal of Portfolio Management* **10**: 68–75.
Sullivan R, Timmermann A, White H. 1999. Data-snooping, technical trading rule performance, and the bootstrap. *Journal of Finance* **54**: 1647–1691.
White H. 2000. A reality check for data snooping. *Econometrica* **68**: 1097–1126.

Authors' biographies:

Dmitry Danilov is currently a postdoctoral fellow in the research group Statistical Information and Modelling at Eurandom, Eindhoven University of Technology. He graduated in mathematics from St. Petersburg State University. From 1997 to 2002 he was a PhD student at Tilburg University, where he defended his thesis *The effects of pretesting in econometrics with applications in finance* in February 2003.

Jan R. Magnus is Research Professor of Econometrics at Tilburg University in The Netherlands. He studied econometrics at the University of Amsterdam, obtaining his MSc (cum laude) in 1975 and his PhD in 1981. His principal affiliations have been the University of British Columbia (1979–1981), the London School of Economics (1981–1996) and CentER, Tilburg University (since 1988, from 1996 as Research Professor). Magnus is a Fellow of the *Journal of Econometrics* (since 1995) and winner of the Econometric Theory Award (1997).

Authors' addresses:

Dmitry Danilov, Eurandom, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

Jan R. Magnus, CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands.

