Tilburg University

Forecast Accuracy after Pretesting with an Application to the Stock Market

Danilov, D.L.; Magnus, J.R.

Publication date:
2002

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):

Danilov, D. L., & Magnus, J. R. (2002). *Forecast Accuracy after Pretesting with an Application to the Stock Market*. (Center Discussion Paper; Vol. 2002-76). Econometrics.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

No. 2002-76

**FORECAST ACCURACY AFTER PRETESTING WITH
AN APPLICATION TO THE STOCK MARKET**

By Dmitry Danilov, Jan R. Magnus

August 2002

ISSN 0924-7815

Discussion paper

Forecast accuracy after pretesting with an application to the stock market

Dmitry Danilov and Jan R. Magnus
CentER, Tilburg University

August 2002

Affiliation: CentER, Tilburg University
P.O. Box 90153
5000 LE Tilburg
The Netherlands

phone: +31-13-466-3092
fax: +31-13-466-3066
e-mail: magnus@kub.nl

Title: Forecast accuracy after pretesting with an application to the stock market

Proposed running head: Forecast accuracy after pretesting

Mailing address:

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

Keywords: Pretest, Forecasting, Model selection, Stock returns.

JEL Codes: C20, C53, G10

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 very substantial. We obtain explicit expressions for this error. To illustrate the theory we consider the regression approach of Pesaran and Timmermann (1994) to stock market forecasting, and show that their proposed recursive predictions are much less robust than naive econometrics might suggest.

1 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 $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \gamma\mathbf{z} + \boldsymbol{\varepsilon}$, where we are uncertain whether to include $\mathbf{z}$ 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.

Section 2 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 Section 3 (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 Section 4 we apply the theory to the problem of forecasting stock market moves (Pesaran and Timmermann, 1994, 1995), and show that the recommendations of Pesaran and Timmermann are much less robust than naive econometrics would seem to imply, thus questioning the usefulness of the implied switching-portfolio strategy. In Section 5 we present a continuous analogue of pretesting which can greatly improve the properties of forecasts. In Section 6 we address the problem of how to incorporate the (obvious) fact that σ^2 is not known in our theory and applications. The effect of this extension is small. Some conclusions are offered in Section 7.

2 Set-up, notation, and preliminary results

The set-up is the same as in Magnus and Durbin (1999) and Danilov and Magnus (2001). We consider the standard linear regression model

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

where $\mathbf{y}$ ($n \times 1$) is the vector of observations, $\mathbf{X}$ ($n \times k$) and $\mathbf{Z}$ ($n \times m$) 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 $(\mathbf{X} : \mathbf{Z})$ has full column-rank $k + m$, and that the disturbances $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ are i.i.d. $N(0, \sigma^2)$.²

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

We define the matrices

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

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

Let $\mathbf{S}_i$ be an $m \times r_i$ selection matrix of rank r_i ($0 \leq r_i \leq m$), so that $\mathbf{S}_i' = (\mathbf{I}_{r_i} : \mathbf{O})$ or a column-permutation thereof. The equation $\mathbf{S}_i'\boldsymbol{\gamma} = \mathbf{0}$ thus selects a subset of the γ 's to be equal to zero. Following Danilov and Magnus (2001), the LS estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ under the restriction $\mathbf{S}_i'\boldsymbol{\gamma} = \mathbf{0}$ are then given by

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

where

$$\mathbf{W}_i = \mathbf{I}_m - \mathbf{P}_i, \quad \mathbf{P}_i = (\mathbf{Z}'\mathbf{M}\mathbf{Z})^{-1/2}\mathbf{S}_i(\mathbf{S}_i'(\mathbf{Z}'\mathbf{M}\mathbf{Z})^{-1}\mathbf{S}_i)^{-1}\mathbf{S}_i'(\mathbf{Z}'\mathbf{M}\mathbf{Z})^{-1/2}$$

are symmetric idempotent $m \times m$ matrices of ranks $m - r_i$ and r_i respectively. (If $r_i = 0$ then $\mathbf{P}_i = \mathbf{O}$.) The distribution of $\mathbf{b}_{(i)}$ is given by

$$\mathbf{b}_{(i)} \sim N(\boldsymbol{\beta} + \sigma\mathbf{Q}\mathbf{P}_i\boldsymbol{\eta}, \sigma^2((\mathbf{X}'\mathbf{X})^{-1} + \mathbf{Q}\mathbf{W}_i\mathbf{Q})). \quad (2)$$

There are 2^m different models to consider, one for each subset of $\gamma_1, \dots, \gamma_m$ 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

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

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

that the model selection is based exclusively on the residuals from the unrestricted model, that is, on $\mathbf{M}\mathbf{y}$. This assumption appears to be satisfied in all standard cases. (Note that the residuals in the i -th model can always be expressed as $\mathbf{e}_{(i)} = \mathbf{D}_i\mathbf{M}\mathbf{y}$ for some idempotent matrix $\mathbf{D}_i$.) More generally, a WALS (weighted-average least-squares) estimator of $\boldsymbol{\beta}$ is defined as $\mathbf{b} = \sum_i \lambda_i \mathbf{b}_{(i)}$, where the weights satisfy $\lambda_i = \lambda_i(\mathbf{M}\mathbf{y})$, $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$, and the sum is taken over all 2^m models. Clearly, the pretest estimator is a special case of the WALS estimator when all λ_i 's are 0 except one which is 1.

The WALS estimator can be written as $\mathbf{b} = \mathbf{b}_r - \mathbf{Q}\mathbf{W}\hat{\boldsymbol{\theta}}$, where $\mathbf{W} = \mathbf{I}_m - \mathbf{P}$ and $\mathbf{P} = \sum_i \lambda_i \mathbf{P}_i$. (Notice that both $\mathbf{P}$ and $\mathbf{W}$ are random matrices, because the $\{\lambda_i\}$ are random.) The equivalence theorem (for estimation) now says that

$$\mathbf{E}(\mathbf{b}) = \boldsymbol{\beta} - \sigma \mathbf{Q} \mathbf{E}(\mathbf{W}\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}), \quad \text{var}(\mathbf{b}) = \sigma^2 \left((\mathbf{X}'\mathbf{X})^{-1} + \mathbf{Q} \text{var}(\mathbf{W}\hat{\boldsymbol{\eta}}) \mathbf{Q}' \right),$$

and hence that

$$\text{MSE}(\mathbf{b}) = \sigma^2 \left((\mathbf{X}'\mathbf{X})^{-1} + \mathbf{Q} \text{MSE}(\mathbf{W}\hat{\boldsymbol{\eta}}) \mathbf{Q}' \right),$$

showing that the properties of the complicated WALS (pretest) estimator $\mathbf{b}$ of $\boldsymbol{\beta}$ depend critically on the properties of the less complicated estimator $\mathbf{W}\hat{\boldsymbol{\eta}}$ of $\boldsymbol{\eta}$.

3 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 $\mathbf{S}'_i \boldsymbol{\gamma} = \mathbf{0}$ the one-period-ahead LS forecast is given by

$$\begin{aligned} \hat{y}_{n+1}^{(i)} &= \mathbf{x}'_{n+1} \mathbf{b}_{(i)} + \mathbf{z}'_{n+1} \mathbf{c}_{(i)} \\ &= \mathbf{x}'_{n+1} \left(\mathbf{b}_r - \mathbf{Q}\mathbf{W}_i \hat{\boldsymbol{\theta}} \right) + \mathbf{z}'_{n+1} \left((\mathbf{Z}'\mathbf{M}\mathbf{Z})^{-1/2} \mathbf{W}_i \hat{\boldsymbol{\theta}} \right) \\ &= \mathbf{x}'_{n+1} \mathbf{b}_r - \boldsymbol{\omega}' \mathbf{W}_i \hat{\boldsymbol{\theta}} = \mathbf{x}'_{n+1} \mathbf{b}_r - \sigma \boldsymbol{\omega}' \mathbf{W}_i \hat{\boldsymbol{\eta}}, \end{aligned}$$

where

$$\boldsymbol{\omega} = \mathbf{Q}' \mathbf{x}_{n+1} - (\mathbf{Z}'\mathbf{M}\mathbf{Z})^{-1/2} \mathbf{z}_{n+1},$$

and $\mathbf{x}_{n+1}$ and $\mathbf{z}_{n+1}$ 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

$$\hat{y}_{n+1} = \sum_i \lambda_i \hat{y}_{n+1}^{(i)} = \mathbf{x}'_{n+1} \mathbf{b}_r - \sigma \boldsymbol{\omega}' \mathbf{W} \hat{\boldsymbol{\eta}}. \quad (3)$$

Notice that $\sigma \hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\theta}}$ and can thus be observed, but that nevertheless $\hat{y}_{n+1}$ depends on σ , because $\mathbf{W}$ (through λ_i) depends on σ .

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

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

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:

$$\text{E}(\text{FE}) = -\sigma \boldsymbol{\omega}' \text{E}(\mathbf{W} \hat{\boldsymbol{\eta}} - \boldsymbol{\eta}),$$

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

and hence

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

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

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 $\boldsymbol{\eta}$ and σ^2 , but not on $\boldsymbol{\beta}$. Secondly, it helps us to find an *optimal* forecast. If we can find λ_i 's such that $\mathbf{W} \hat{\boldsymbol{\eta}}$ is an optimal estimator of $\boldsymbol{\eta}$ (in the sense of minimizing the mean squared error), 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 $\boldsymbol{\eta}$ was studied in Magnus (2002), and led to the ‘neutral Laplace’ estimator. In Section 5 we shall apply the Laplace weights to forecasting, and demonstrate the superiority 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, the reported MSFE (variance), denoted $\widetilde{\text{MSFE}}$, is given by

$$\widetilde{\text{MSFE}} = \sigma^2 (\mathbf{x}'_{n+1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{n+1} + \boldsymbol{\omega}'\mathbf{W}\boldsymbol{\omega} + 1), \quad (4)$$

and the reported 95% prediction interval for y_{n+1} is

$$\hat{y}_{n+1} \pm 1.96\sigma\sqrt{\mathbf{x}'_{n+1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{n+1} + \boldsymbol{\omega}'\mathbf{W}\boldsymbol{\omega} + 1}, \quad (5)$$

where σ is estimated by some consistent estimator $\hat{\sigma}$. 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

$$\psi_1(\boldsymbol{\eta}) := \boldsymbol{\omega}'\text{E}(\mathbf{W}\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \quad (6)$$

and

$$\psi_2(\boldsymbol{\eta}) := \mathbf{x}'_{n+1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{n+1} + \boldsymbol{\omega}'\text{var}(\mathbf{W}\hat{\boldsymbol{\eta}})\boldsymbol{\omega}, \quad (7)$$

both of which depend also on σ , because $\mathbf{W}$ depends on σ . Then, by Theorem 1,

$$\text{FE} \sim (-\sigma\psi_1(\boldsymbol{\eta}), \sigma^2(\psi_2(\boldsymbol{\eta}) + 1)),$$

so that an approximate 95% prediction interval for y_{n+1} is given by

$$\hat{y}_{n+1} + \sigma \left(\psi_1(\boldsymbol{\eta}) \pm 1.96\sqrt{\psi_2(\boldsymbol{\eta}) + 1} \right). \quad (8)$$

The interval is approximate because the distribution of FE is not normal. Furthermore, in contrast to (5), the interval depends on $\boldsymbol{\eta}$ (and on σ of course), which is unknown. We obtain an estimated prediction interval by replacing $\boldsymbol{\eta}$ and σ by the estimates $\hat{\boldsymbol{\eta}}$ and $\hat{\sigma}$.

When the number of observations n becomes large, then $\hat{\sigma}$ will converge to σ , but $\hat{\boldsymbol{\eta}}$ will not converge to $\boldsymbol{\eta}$, because $\text{var}(\hat{\boldsymbol{\eta}}) = \mathbf{I}_m$. Hence, $\hat{\boldsymbol{\eta}}$ is an unbiased but not a consistent estimator of $\boldsymbol{\eta}$. To protect ourselves against ‘large’ deviations of $\hat{\boldsymbol{\eta}}$ from $\boldsymbol{\eta}$, we shall also consider the more conservative interval

$$\hat{y}_{n+1} + \sigma C_1(\hat{\boldsymbol{\eta}}) < y_{n+1} < \hat{y}_{n+1} + \sigma C_2(\hat{\boldsymbol{\eta}}), \quad (9)$$

where

$$C_1(\hat{\boldsymbol{\eta}}) := \min_{\boldsymbol{\eta} \in \mathcal{H}(\hat{\boldsymbol{\eta}})} \left(\psi_1(\boldsymbol{\eta}) - 1.96\sqrt{\psi_2(\boldsymbol{\eta}) + 1} \right)$$

and

$$C_2(\hat{\boldsymbol{\eta}}) := \max_{\boldsymbol{\eta} \in \mathcal{H}(\hat{\boldsymbol{\eta}})} \left(\psi_1(\boldsymbol{\eta}) + 1.96 \sqrt{\psi_2(\boldsymbol{\eta}) + 1} \right).$$

The set $\mathcal{H}$ is an m -dimensional cube, defined by $\mathcal{H}(\hat{\boldsymbol{\eta}}) := \{\boldsymbol{\eta} : |\hat{\eta}_i - \eta_i| < a_m, i = 1, \dots, m\}$, where a_m is determined such that, for standard-normal u , $\Pr(|u| < a_m)^m = 0.95$. (In our application, $m = 4$ and hence $a_m = 2.49$.)

4 Forecasting stock returns

In order to investigate the effects of ignoring pretesting on forecasts in practice, we will consider a question from the finance literature. Perhaps the first important application of linear regression in finance is the capital asset pricing model. Black, Jensen and Scholes (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. Equity risk premia related variables, such as the dividend yield, were suggested by Rozeff (1984), while French, Schwert and Stambaugh (1987) proposed default bond premia. Fama and French (1989) suggested to use the interest rate as an explanatory variable, since it affects the overall economic activity and, as a consequence, the stock market activity. Using the inflation rate (or an inflation-related characteristic) as an explanatory variable goes back to Lucas (1976). Industrial production variables were used by Chen, Roll and Ross (1986) and Balvers, Cosimano and McDonald (1990). A price-earnings variable, describing the relationship of the stock price and the actual earnings of the company, was used in Fama and French (1992). Inspired by the development of regression models, Cheng, Lo and Ma (1990) attempted to forecast the Hong Kong stock price index. Their regression models were however not sufficiently powerful to effectively predict the direction of the change in the index. Pesaran and Timmermann (1994) were more successful and 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 shall reconsider the question discussed by Pesaran and Timmermann

(1994), hereafter PT94: can the annual excess returns on common stocks for the Standard & Poor 500 (SP 500) index be predicted?³ Of course, PT94 pretested. In fact, they state explicitly (p. 339) that they “experimented with a number of specifications”. The dependent variable in the linear regression is ρ_t , the excess returns in year t . In analyzing the effect of pretesting we have to decide which regressors play a role and which of these are focus regressors and which are auxiliary. The distinction is not completely unambiguous, but we decided — after carefully studying their model selection description — that PT94’s model contains four focus regressors ($k = 4$) and four auxiliary regressors ($m = 4$).⁴ The focus regressors are:

constant term,

PI_{t-2} : annual inflation rate (lagged two periods),

$DI3_{t-1}$: change in 3-month T-bill rate (lagged one period),

$TERM_{t-1}$: term premium (lagged one period),

and 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).

Employing a forward (specific-to-general) model selection procedure, PT94 (p. 339) then obtain the following estimated model of the annual excess returns over the period 1954–1991:

$$\hat{\rho}_t = -0.289 - 1.72 PI_{t-2} - 0.06 DI3_{t-1} + 0.11 TERM_{t-1} + 9.17 YSP_{t-1}.$$

(0.077) (0.44) (0.02) (0.04) (2.02)

We could not acquire *exactly* the same data set as PT94, but we almost could. In addition, since our data set extends to the year 2001, we had to employ a slightly different definition of the term premium $TERM_{t-1}$.⁵ Our

³PT94 also consider the Dow Jones Industrial portfolio, and also monthly and quarterly frequencies. We shall only consider the SP 500 index and annual returns.

⁴In fact, PT94 did more pretesting than we analyze in this paper, so that $m > 4$ and the effect of ignoring pretesting is even larger than we report.

⁵PT94 measure the term premium as the difference between the 6-month commercial paper rate (risky) and the 3-month T-bill rate (riskless) in January. Since the 6-month commercial paper rate does not exist after 1997, we use the 3-month financial paper rate instead.

data set thus contains eight annual time series (plus a constant term) over 46 years (1956–2001).⁶ A full description and all the data are given in the appendix.

With our data set we re-estimated the annual excess returns over the period 1956–1991, also employing a forward pretest procedure. This led to the same model as obtained by PT94, but to slightly different estimates:

$$\hat{\rho}_t = -0.343 - 1.65 \text{PI}_{t-2} - 0.04 \text{DI3}_{t-1} + 0.17 \text{TERM}_{t-1} + 10.14 \text{YSP}_{t-1}.$$

(0.084) (0.44) (0.02) (0.04) (2.17)

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 $\mathbf{X}$ -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 an well-defined and unambiguous manner. Notice however that in the final model there is no guarantee that all t -statistics are significant.

Secondly, 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 takes to be 1.96. This does not make any serious difference, and is more in line with the normality assumptions made in the approximations.

Thirdly, all computations are performed by Monte-Carlo methods, based on 1,000 replications, and properly tested for stability.

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.

⁶We could not obtain the full data set for 1954 and 1955, because TERM and YSP are not available in 1953 and 1954.

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

In Figure 1, the solid line gives the one-period-ahead forecasts $\hat{y}_{n+1}$, while the little open circles give the realized values y_{n+1} . The forecasts are the same, whether we take pretesting into account or not. The difference lies in the prediction bounds. 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 as would have been reported by PS94. 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.

The two dash-dotted lines show the approximate 95% prediction bounds of the pretest forecast, based on (8), while the dashed lines give the more conservative interval, based on (9). Because of the bias effect, these 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.

In doing the calculations for the dash-dotted and the dashed intervals, we estimate σ^2 by the LS estimator in the unrestricted model, that is,

$$\hat{\sigma}^2 = \frac{1}{n - k - m} (\mathbf{y} - \mathbf{X}\mathbf{b}_u - \mathbf{Z}\hat{\gamma})'(\mathbf{y} - \mathbf{X}\mathbf{b}_u - \mathbf{Z}\hat{\gamma}), \quad (10)$$

which simplifies the calculations without effecting the results; see Section 6.⁷

Although the WALS 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, 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 within the restrictions of the PS94 data set.

Lack of sensitivity in one direction does not, however, imply lack of sensitivity in another direction. For the question posed in PS94, the most important estimate is not the forecast, but rather — as argued by Granger and Pesaran (2000) — 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

$$\begin{aligned}\Pr(y_{n+1} > 0) &= \Pr(\mathbf{x}'_{n+1}\boldsymbol{\beta} + \mathbf{z}'_{n+1}\boldsymbol{\gamma} + \varepsilon_{n+1} > 0) \\ &= \Pr(-\varepsilon_{n+1} < \mathbf{x}'_{n+1}\boldsymbol{\beta} + \mathbf{z}'_{n+1}\boldsymbol{\gamma}) \\ &= \Phi\left(\frac{\mathbf{x}'_{n+1}\boldsymbol{\beta} + \mathbf{z}'_{n+1}\boldsymbol{\gamma}}{\sigma}\right),\end{aligned}\tag{11}$$

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 market will go up 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 the market will go down and will invest in bonds. Of course, the probability that $\Pr(y_{n+1} > 0)$ is not known and needs to be estimated by $\Phi(\hat{y}_{n+1}/\hat{\sigma})$. Moreover, we want to know how good the estimates are, using appropriate prediction intervals.

⁷Notice that ψ_1 and ψ_2 , and therefore C_1 and C_2 , also depend on σ .

Figure 2. Pretest forecast probabilities $\Pr(y_{n+1} > 0)$ with three sets of prediction intervals.

The solid line in Figure 2 gives the estimated probability that $\Pr(y_{n+1} > 0)$. If we take no account of the effects of pretesting, then a 95% prediction interval for the parameter $(\mathbf{x}'_{n+1}\boldsymbol{\beta} + \mathbf{z}'_{n+1}\boldsymbol{\gamma})/\sigma$ is given by

$$\frac{\hat{y}_{n+1}}{\sigma} \pm 1.96\sqrt{\mathbf{x}'_{n+1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{n+1} + \boldsymbol{\omega}'\mathbf{W}\boldsymbol{\omega}}, \quad (12)$$

and the dotted lines are based on these bounds.

If, however, we do take account of pretesting, then

$$\hat{y}_{n+1} \sim (\mathbf{x}'_{n+1}\boldsymbol{\beta} + \mathbf{z}'_{n+1}\boldsymbol{\gamma} - \sigma\psi_1(\boldsymbol{\eta}), \sigma^2\psi_2(\boldsymbol{\eta})), \quad (13)$$

where ψ_1 and ψ_2 are defined in (6) and (7), so that an approximate 95% prediction interval for $(\mathbf{x}'_{n+1}\boldsymbol{\beta} + \mathbf{z}'_{n+1}\boldsymbol{\gamma})/\sigma$ is given by

$$\frac{\hat{y}_{n+1}}{\sigma} + \psi_1(\boldsymbol{\eta}) \pm 1.96\sqrt{\psi_2(\boldsymbol{\eta})}. \quad (14)$$

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

Finally, as in Section 3, we obtain more conservative bounds, taking into account that $\hat{\boldsymbol{\eta}}$, although unbiased, is inconsistent:

$$\frac{\hat{y}_{n+1}}{\sigma} + C_3(\hat{\boldsymbol{\eta}}) < \frac{\mathbf{x}'_{n+1}\boldsymbol{\beta} + \mathbf{z}'_{n+1}\boldsymbol{\gamma}}{\sigma} < \frac{\hat{y}_{n+1}}{\sigma} + C_4(\hat{\boldsymbol{\eta}}), \quad (15)$$

where

$$C_3(\hat{\boldsymbol{\eta}}) := \min_{\boldsymbol{\eta} \in \mathcal{H}(\hat{\boldsymbol{\eta}})} \left(\psi_1(\boldsymbol{\eta}) - 1.96\sqrt{\psi_2(\boldsymbol{\eta})} \right),$$

and

$$C_4(\hat{\boldsymbol{\eta}}) := \max_{\boldsymbol{\eta} \in \mathcal{H}(\hat{\boldsymbol{\eta}})} \left(\psi_1(\boldsymbol{\eta}) + 1.96\sqrt{\psi_2(\boldsymbol{\eta})} \right).$$

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 (correct) pretest prediction intervals are such that we cannot be confident in *any* year. This is true for the dash-dotted lines and, *a fortiori*, for the more conservative dashed lines.

The difference between the dotted and the dash-dotted lines is twice as large as in Figure 1, on average 2.6 times as wide. This large difference in the effects of ignoring pretesting between Figures 1 and 2 (the dash-dotted lines) can be attributed completely to a small difference between formulas (8) and (14). The first formula contains the term $\psi(\boldsymbol{\eta}) + 1$, which is replaced by $\psi(\boldsymbol{\eta})$ in the second formula. The simple appearance of one 1 thus appears to have a large effect on the bounds.

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 of forecasting are thus serious and should be analyzed and incorporated in econometric analyses.

5 Optimal forecasts using the 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. But which weights

should be taken? In Magnus (2002) a Bayesian solution to this problem is proposed (in the estimation context) for the case $m = 1$. 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 $\lambda = \lambda(\hat{\eta})$ is

$$\lambda(\hat{\eta}) = \frac{\int \eta \pi(\eta) \exp(-(\hat{\eta} - \eta)^2/2) d\eta}{\hat{\eta} \int \pi(\eta) \exp(-(\hat{\eta} - \eta)^2/2) d\eta},$$

where the prior π is the ‘neutral’ Laplace density,

$$\pi(\eta) = \frac{c}{2} \exp(-c|\eta|), \quad -\infty < \eta < \infty, \quad c = \log 2.$$

The neutrality of the prior guarantees that $\text{median}(\eta) = 0$ and $\text{median}(\eta^2) = 1$. We know that the use of the Laplace weights leads to better estimates (admissible to begin with) than the pretest weights.

When $m > 1$ it is not so clear how the weights should be taken. However, in the special case where $\mathbf{Z}'\mathbf{M}\mathbf{Z} = \mathbf{I}_m$, the multi-dimensional problem separates into m one-dimensional problems, and we can use the Laplace weights for each dimension separately; see Danilov and Magnus (2001, Theorem 3).

Let us consider the ‘orthogonalization’ $\mathbf{Z}'\mathbf{M}\mathbf{Z} = \mathbf{I}_m$ in some more detail. Orthogonalization can always be achieved by taking appropriate linear combinations of the m auxiliary regressors in $\mathbf{Z}$ (leaving the focus regressors unchanged). More specifically, let $\mathbf{T}_1$ be an orthogonal $m \times m$ matrix such that $\mathbf{T}_1' \mathbf{Z}' \mathbf{M} \mathbf{Z} \mathbf{T}_1 = \mathbf{A}$ (diagonal). Then, letting $\mathbf{T} = \mathbf{T}_1 \mathbf{A}^{-1/2}$, we have $\mathbf{T}' \mathbf{Z}' \mathbf{M} \mathbf{Z} \mathbf{T} = \mathbf{I}_m$. Now define new auxiliary regressors $\mathbf{Z}^* = \mathbf{Z} \mathbf{T}$ and $\mathbf{z}_{n+1}^* = \mathbf{T}' \mathbf{z}_{n+1}$. Then, clearly, $\mathbf{Z}^{*'} \mathbf{M} \mathbf{Z}^* = \mathbf{I}_m$. As a consequence of this transformation, $\boldsymbol{\omega}$, $\mathbf{R}(\boldsymbol{\eta}) := \text{MSE}(\mathbf{W}\hat{\boldsymbol{\eta}})$, and MSFE will all change, but $\boldsymbol{\omega}'\boldsymbol{\omega}$ will not change. This follows because

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

so that

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

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

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

and

$$\text{MSFE}^* = \sigma^2(\boldsymbol{\omega}^{*\prime}\boldsymbol{\omega}^*) \left(\frac{\mathbf{x}'_{n+1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{n+1} + 1}{\boldsymbol{\omega}^{*\prime}\boldsymbol{\omega}^*} + \frac{\boldsymbol{\omega}^{*\prime}\mathbf{R}^*(\boldsymbol{\eta})\boldsymbol{\omega}^*}{\boldsymbol{\omega}^{*\prime}\boldsymbol{\omega}^*} \right)$$

lies in the two expressions

$$\xi^2 := \frac{\boldsymbol{\omega}'\mathbf{R}(\boldsymbol{\eta})\boldsymbol{\omega}}{\boldsymbol{\omega}'\boldsymbol{\omega}} \quad \text{and} \quad \xi^{*2} := \frac{\boldsymbol{\omega}^{*\prime}\mathbf{R}^*(\boldsymbol{\eta})\boldsymbol{\omega}^*}{\boldsymbol{\omega}^{*\prime}\boldsymbol{\omega}^*}.$$

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. Secondly, while the eigenvalues of $\mathbf{R}(\boldsymbol{\eta})$ are not necessarily bounded, the eigenvalues of $\mathbf{R}^*(\boldsymbol{\eta})$ are always bounded, so that ξ^{*2} is always finite even when ξ^2 is infinite.⁸ Thirdly, 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 three procedures: forward, orthogonal, and Laplace. The forward pretest procedure was already discussed and applied; it is the standard procedure used in applied work. The orthogonal procedure first transforms the auxiliary regressors $\mathbf{Z}$ so that they become ‘orthogonal’ (in the sense that $\mathbf{Z}'\mathbf{M}\mathbf{Z} = \mathbf{I}_m$), and then applies the standard pretest procedure to the transformed model. In this case it does not matter whether the pretest procedure is forward, backward, or something else; they all lead to the same result. Finally, the Laplace procedure is based on the transformed model, but it will use *all* auxiliary (transformed) regressors. The weights λ_i will 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.

⁸In the forward pretest procedure, ξ^2 can become as large as we please by making $\mathbf{M}\mathbf{z}_i$ and $\mathbf{M}\mathbf{z}_j$ more and more correlated; see Danilov and Magnus (2001, Section 7).

Figure 3. Forecast probabilities $\Pr(y_{n+1} > 0)$ for three procedures.

The main conclusion from Figure 3 is that none of the three procedures considered predict particularly well. The 2001 crash, for example, was only predicted by the Laplace procedure. The triangles depict the direction of the market: down in 1995 and 2001, up in the other eight years. Of the thirty predictions (10 years, 3 procedures), exactly one half were correct. For example, in 1992, pretest and Laplace predicted correctly, but orthogonal predicted incorrectly. In 1996 and 1998 the market went up, but all three procedures predicted that it would go down. It turns out that predicting the annual excess returns on common stocks is a very hazardous business.

This does not, however, imply that all three *procedures* are equally bad. Let us consider the MSFE of the WALS estimator, given in Theorem 1:

$$\begin{aligned} \text{MSFE} &= \sigma^2 (\mathbf{x}'_{n+1} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_{n+1} + \boldsymbol{\omega}' \text{MSE}(\mathbf{W} \hat{\boldsymbol{\eta}}) \boldsymbol{\omega} + 1) \\ &= \sigma^2 (\psi_2(\boldsymbol{\eta}) + \psi_1^2(\boldsymbol{\eta}) + 1). \end{aligned}$$

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

$$\text{MSFE} < \sigma^2 (C_5(\hat{\boldsymbol{\eta}}) + 1),$$

where

$$C_5(\hat{\boldsymbol{\eta}}) := \max_{\boldsymbol{\eta} \in \mathcal{H}(\hat{\boldsymbol{\eta}})} (\psi_2(\boldsymbol{\eta}) + \psi_1^2(\boldsymbol{\eta})).$$

In Figure 4 we compare the bounds of the MSFE for the forward pretest, orthogonal, and Laplace procedures.

Figure 4. Upper bounds of MSFE, σ^2 estimated by (10).

Figure 4 shows convincingly the superiority of the Laplace estimator. Its MSFE bound is very much lower than for the pretest estimator, and uniformly so. Moreover, if we compare the bounds with their theoretical minimum

$$\sigma^2 (\mathbf{x}'_{n+1} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_{n+1} + 1),$$

then the difference between the procedures becomes even more pronounced. We also observe that the MSFE bounds vary significantly over time.

We thus conclude that — if our focus is forecasting rather than model selection — substantially better forecasts can be generated using the Laplace weights.

6 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 (10), 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 sub-model

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

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 $\mathbf{M}\mathbf{y}$, so that Theorem 1 still applies, but they are much more complicated and time-consuming.

Figure 5. Upper bounds of MSFE, σ^2 estimated ‘properly’.

In Figure 5 we recalculate the MSFE-bounds of Figure 4, but now taking the estimation of σ^2 into proper account. As the plots show, the difference between Figures 4 and 5 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.

7 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 (or become) 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. The conclusions of PT94 are much less robust than naive econometrics might imply, when the effects of pretesting are properly accounted for.

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

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. Secondly, our data set extends to the year 2001, so that we had to employ a slightly different definition of the term premium TERM, since the 6-month commercial paper rate is no longer published by the Federal Reserve. Thirdly, 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 1 and 2 provide the full data set employed.

dependent variable

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

$$\rho_t = \text{NRSP}_t - \text{I12}_{t-1},$$

where

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

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.⁹

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 $DIVSP = PSP \times YSP$, 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 U.S. Department of Labor, Bureau of Labor Statistics, Series: Producer Price Index by Finished Goods (April 1947 to present).¹⁰

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.¹¹

The fourth focus regressor is TERM, the term premium, defined as the difference between the 3-month financial paper rate (IF3:JAN) and I3:JAN. PT94 measure the term premium as the difference between the 6-month commercial paper rate (risky) and the 3-month T-bill rate (riskless) 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,

⁹See <http://www.federalreserve.gov/releases/h15/data/b/tbsm1y.txt>.

¹⁰Available online at www.bls.gov.

¹¹See <http://www.federalreserve.gov/Releases/h15/data/b/tbsm3m.txt>.

Monthly (1955–1997), and H15 Federal Reserve Statistical Release, section Weekly Releases, Selected Interest Rates, Historical data, Commercial paper (Financial), 3-month, Monthly (1997–2002).¹²

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).

Secondly, 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-line database of the Federal Reserve Bank of St.-Louis.¹³ The data are monthly, seasonally adjusted, and range from January 1940 to August 2001. The data series is an index, base year 1992.

Thirdly, 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 by BCI Data Manager (January 1948 to November 1995).¹⁴ For more recent data we extend the series as follows. We take the ‘updated series’ from the Economagic website.¹⁵ 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.

¹²See <http://www.federalreserve.gov/Releases/h15/data/m/hfp3m.txt> for the historical data and [.../fp3m.txt](http://www.federalreserve.gov/Releases/h15/data/m/hfp3m.txt) for the recent data.

¹³See www.stls.frb.org, which in turn refers to the Federal Reserve Board, Washington, D.C.

¹⁴See <http://www.wfu.edu/~cottrell/bci/Software.html>.

¹⁵See <http://www.economagic.com/em-cgi/data.exe/feddal/jlead>.

<i>year</i>	ρ_t	PI_{t-2}	$DI3_{t-1}$	$TERM_{t-1}$
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

Table 1a. Dependent variable and focus regressors, 1956–1991.

<i>year</i>	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

Table 2a. Auxiliary regressors, 1956–1991.

<i>year</i>	ρ_t	PI_{t-2}	$DI3_{t-1}$	$TERM_{t-1}$
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

Table 1b. Dependent variable and focus regressors, 1992–2001.

<i>year</i>	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

Table 2b. Auxiliary regressors, 1992–2001.

References

- Abadir, K.M. and J.R. Magnus (2002). Notation in econometrics: a proposal for a standard, *The Econometrics Journal*, 5, 76–90.
- Balvers, R.J., T.F. Cosimano, and B. McDonald (1990). Predicting stock returns in an efficient market, *Journal of Finance*, 55, 1109–1128.
- Black, F., M.C. Jensen, and M. Scholes (1972). The capital asset pricing model: some empirical tests, in: M.C. Jensen (ed.), *Studies in the Theory of Capital Markets*, New York: Praeger, 79–121.
- Chen, N.F., R. Roll, and S.A. Ross (1986). Economic forces and the stock market, *Journal of Business*, 59, 382–403.
- Cheng, T.C.E., Y.K. Lo and K.W. Ma (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.
- Danilov, D. and J.R. Magnus (2001). On the harm that pretesting does, CentER Discussion Paper, Tilburg University, The Netherlands.
- Fama, E.F. and K.R. French (1989). Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics*, 25, 23–49.
- Fama, E.F. and K.R. French (1992). The cross-section of expected stock returns, *Journal of Finance*, 47, 427–465.
- Fama, E.F. and J.D. MacBeth (1973). Risk, return and equilibrium: empirical tests, *Journal of Political Economy*, 81, 607–636.
- French, K.R., G.W. Schwert, and R. Stambaugh (1987). Expected stock returns and volatility, *Journal of Financial Economics*, 19, 3–30.
- Granger, C.W.J. and M.H. Pesaran (2000). Economic and statistical measures of forecast accuracy, *Journal of Forecasting*, 19, 537–560.
- Lucas, R.E., Jr. (1976). Econometric policy evaluation: a critique, in: K. Brunner and A.H. Meltzer (eds.), *The Phillips Curve and Labor Markets*, Carnegie-Rochester Conference Series on Public Policy, Vol. 1, Amsterdam: North-Holland, 19–46.

- Magnus, J.R. (2002). Estimation of the mean of a univariate normal distribution with known variance, *The Econometrics Journal*, 5, 225–236.
- 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.
- Pesaran, M.H. and A. Timmermann (1994). Forecasting stock returns. An examination of stock market trading in the presence of transaction costs, *Journal of Forecasting*, 13, 335–367.
- Pesaran, M.H. and A. Timmermann (1995). Predictability of stock returns: robustness and economic significance, *Journal of Finance*, 50, 1201–1228.
- Rozeff, M.S. (1984). Dividend yields are equity risk premiums, *Journal of Portfolio Management*, 10, 68–75.