Forecast accuracy after pretesting with an application to the stock market[*]

Dmitri L. Danilov and Jan R. Magnus
CentER, Tilburg University

13 February 2002

Version 1 — Very Preliminary
Prepared for presentation at ESEM 2002 in Venice
NOT TO BE QUOTED

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

---

[*] We are grateful to seminar participants at X, Y, and Z for constructive and useful comments. Correspondence to: Jan R. Magnus, CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands, 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: As a rule, econometric data are non-experimental. The same data set is used to select the model and, conditional on the selected model, to forecast. In applied econometrics, 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

Econometrics, like astronomy but unlike biology, medicine and physics, is a non-experimental science. Typically we 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 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \gamma\boldsymbol{z} + \boldsymbol{\varepsilon}$, where we are uncertain whether to include $\boldsymbol{z}$ or not. The usual procedure is to compute the t-statistic on γ, and then, depending on whether $|t|$ is 'large' or 'small', decide to use the unrestricted or the restricted 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 mean square forecast error of the pretest forecast. In Section 4 we apply the theory to 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 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

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

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

Let $\boldsymbol{S}_i$ be an $m \times r_i$ selection matrix of rank $r_i \geq 0$, so that $\boldsymbol{S}_i' = (\boldsymbol{I}_{r_i} : \mathbf{O})$ 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 Danilov and Magnus (2001), 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

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

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

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

The WALS (weighted-average least-squares) estimator of $\boldsymbol{\beta}$ is defined as $\boldsymbol{b} = \sum_i \lambda_i \boldsymbol{b}_{(i)}$, where we assume that the weights λ_i satisfy $\lambda_i = \lambda_i(\boldsymbol{M}\boldsymbol{y})$, $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$. These assumptions are satisfied in all standard cases. The sum is taken over all 2^m different models, obtained by setting a subset of $\gamma_1, \ldots, \gamma_m$ equal to zero. The pretest estimator is the special case of the WALS estimator when all λ_i's are 0 except one which is 1.

---

[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.

It follows that $\boldsymbol{b} = \boldsymbol{b}_r - \boldsymbol{Q}\boldsymbol{W}\hat{\boldsymbol{\theta}}$, where $\boldsymbol{W} = \boldsymbol{I}_m - \boldsymbol{P}$ and $\boldsymbol{P} = \sum_i \lambda_i \boldsymbol{P}_i$. (Notice that both $\boldsymbol{P}$ and $\boldsymbol{W}$ are random matrices.)

The equivalence theorem (for estimation) now says that

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

and hence that

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

showing that the properties of the complicated pretest estimator $\boldsymbol{b}$ of $\boldsymbol{\beta}$ depend critically on the properties of the less complicated estimator $\boldsymbol{W}\hat{\boldsymbol{\eta}}$ 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 $\boldsymbol{S}_i'\boldsymbol{\gamma} = \boldsymbol{0}$ the one-period-ahead LS forecast is given by

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

where

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

where $\boldsymbol{x}_{n+1}$ 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. Hence the pretest forecast takes the form

$$\hat{y}_{n+1} = \sum_i \lambda_i \hat{y}_{n+1}^{(i)} = \boldsymbol{x}_{n+1}'\boldsymbol{b}_r - \sigma\boldsymbol{\omega}'\boldsymbol{W}\hat{\boldsymbol{\eta}}. \tag{2}$$

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

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

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

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

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

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

and hence

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

Proof: The essential ingredient is that $\boldsymbol{b}_r$ and $\boldsymbol{My}$ are independent. This implies that $\boldsymbol{b}_r$ and $\boldsymbol{W}\hat{\boldsymbol{\eta}}$ are independent, and hence that $(\boldsymbol{b}_r, \boldsymbol{W}\hat{\boldsymbol{\eta}}, \varepsilon_{n+1})$ are all independent of each other. The results follow. ||

We see that if we can find λ_i's such that $\boldsymbol{W}\hat{\boldsymbol{\eta}}$ 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. Moreover, these λ_i's are also the ones which provide the optimal WALS estimator of $\boldsymbol{\beta}$.

We shall see in Section 5 how the use of optimal λ_i's (based on the 'neutral' Laplace prior) can improve the forecasts.

We now compare the actual MSFE (given in Theorem 1) with the reported MSFE, denoted $\widetilde{\mathrm{MSFE}}$, and given by

$$\widetilde{\mathrm{MSFE}} = \sigma^2 \left(\boldsymbol{x}_{n+1}' (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{x}_{n+1} + \boldsymbol{\omega}' \boldsymbol{W} \boldsymbol{\omega} + 1 \right).$$

To compare MSFE and $\widetilde{\mathrm{MSFE}}$, we define the *underreporting ratio* UR as

$$\mathrm{UR} = 1 - \frac{\widetilde{\mathrm{MSFE}}}{\mathrm{MSFE}} = \frac{\boldsymbol{q}'(\boldsymbol{R}(\boldsymbol{\eta}) - \boldsymbol{W})\boldsymbol{q}}{\boldsymbol{q}'\boldsymbol{R}(\boldsymbol{\eta})\boldsymbol{q} + (1/q_0^2)},$$

where

$$\boldsymbol{R}(\boldsymbol{\eta}) = \mathrm{MSE}(\boldsymbol{W}\hat{\boldsymbol{\eta}}), \qquad \boldsymbol{q} = \frac{\boldsymbol{\omega}}{\sqrt{\boldsymbol{\omega}'\boldsymbol{\omega}}}, \qquad q_0^2 = \frac{\boldsymbol{\omega}'\boldsymbol{\omega}}{\boldsymbol{x}_{n+1}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_{n+1}}.$$

Notice that $\boldsymbol{q}'\boldsymbol{q} = 1$. The UR is a random variable, since it depends on $\boldsymbol{W}$, which depends on $\hat{\boldsymbol{\eta}}$. Both the UR and its expectation are unobservable, since they depend on $\boldsymbol{\eta}$ via $\boldsymbol{R}(\boldsymbol{\eta})$.

We know from Danilov and Magnus (2001, Theorem 3) that major simplifications occur when $\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z} = \boldsymbol{I}_m$. This condition 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 $\boldsymbol{z}_{n+1}^* = \boldsymbol{T}'\boldsymbol{z}_{n+1}$. As a consequence of this transformation, $\boldsymbol{\omega}$, $\boldsymbol{R}(\boldsymbol{\eta})$, and MSFE will all change, but $\boldsymbol{\omega}'\boldsymbol{\omega}$ and q_0^2 will not change. This follows because

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

so that

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

Hence, since $\boldsymbol{T}\boldsymbol{T}' = (\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z})^{-1}$, we see that $\boldsymbol{\omega}^{*\prime}\boldsymbol{\omega}^* = \boldsymbol{\omega}'\boldsymbol{\omega}$, and hence that $q_0^{*2} = q_0^2$. Thus the only difference between

$$\text{MSFE} = \sigma^2(\boldsymbol{\omega}'\boldsymbol{\omega})\left(\frac{1}{q_0^2} + \frac{\boldsymbol{\omega}'\boldsymbol{R}(\boldsymbol{\eta})\boldsymbol{\omega}}{\boldsymbol{\omega}'\boldsymbol{\omega}} + \frac{1}{\boldsymbol{\omega}'\boldsymbol{\omega}}\right),$$

and

$$\text{MSFE}^* = \sigma^2(\boldsymbol{\omega}^{*\prime}\boldsymbol{\omega}^*)\left(\frac{1}{q_0^{*2}} + \frac{\boldsymbol{\omega}^{*\prime}\boldsymbol{R}^*(\boldsymbol{\eta})\boldsymbol{\omega}^*}{\boldsymbol{\omega}^{*\prime}\boldsymbol{\omega}^*} + \frac{1}{\boldsymbol{\omega}^{*\prime}\boldsymbol{\omega}^*}\right),$$

lies in the two expressions

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

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, general-to-specific or specific-to-general), MSFE* is independent of the selection procedure. Secondly, while the eigenvalues of $\boldsymbol{R}(\boldsymbol{\eta})$ are not necessarily bounded, the eigenvalues of $\boldsymbol{R}^*(\boldsymbol{\eta})$ are always bounded, so that ξ^{*2} is always finite even when ξ^2 is infinite. 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 (see Section 5).

4 Forecasting stock returns

In order to investigate the effect of ignoring pretesting on forecasts, we consider 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 reading 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),
$\mathrm{DI3}_{t-1}$: 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 specific-to-general model selection procedure then yields the following estimated model of the annual excess returns over the period 1954–1991:

$$\begin{array}{llllll} \hat{\rho}_t = -\underset{(0.077)}{0.289} - & \underset{(0.44)}{1.72}\,\mathrm{PI}_{t-2} & - \underset{(0.023)}{0.060}\,\mathrm{DI3}_{t-1} & + \underset{(0.040)}{0.11}\,\mathrm{TERM}_{t-1} & + \underset{(2.02)}{9.17}\,\mathrm{YSP}_{t-1}. \end{array} \tag{3}$$

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

---

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

[4] In 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.

to employ a slightly different definition of the term premium TERM_{t-1}.[5] Estimating the annual excess returns over the same period as PT94 and employing a specific-to-general pretest procedure then leads to the same model, but slightly different estimates:

$$\begin{array}{lllllll} \hat{\rho}_t = & -0.398 - & 1.145\,\mathrm{PI}_{t-2} - & 0.0219\,\mathrm{DI3}_{t-1} + & 0.28\,\mathrm{TERM}_{t-1} + & 10.44\,\mathrm{YSP}_{t-1}. & \\ & (0.087) & (0.44) & (0.018) & (0.059) & (2.21) & (4) \end{array}$$

Our data set thus contains eight annual time series (plus a constant term) over 48 years (1954–2001). A full description is given in the appendix.

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 estimated over the period 1954–1999.

We first consider the bias, see Figure 1.

FIGURE 1

The solid line in Figure 1 gives the point forecasts, and the two dotted lines give the standard least-squares 95% confidence bounds (ignoring the effects of pretesting). The two dashed lines show the 95% confidence bounds of the pretest forecast.

A few words of explanation are required here. We know from the equivalence theorem that the bias of the WALS forecast depends on $\boldsymbol{\eta}$ which is not observable. However, we can obtain an interval estimator of the unobservable bias, that is, an interval which covers the true but unknown bias with probability 95%. The dashed lines in Figure 1 represent this interval, centered around the corresponding point forecast.

The confidence interval for the bias turns out to be quite large, in certain points even wider than the standard forecast interval (e.g. year 1999). The WALS forecast is thus seriously biased.

In Figure 2 we consider the variance of the pretest forecast.

FIGURE 2

The solid line and the two dotted lines are the same as in Figure 1. The dashed line now represents the 95% boundary for the standard deviation of

---

[5] 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 WALS forecast. As with the bias, the standard deviation of the WALS forecast is unobservable. But again an interval estimator can be constructed. We see that the standard deviation is seriously underestimated. Using standard forecast intervals is thus very misleading for evaluating the accuracy of the forecast.

Next we consider the total effect of pretesting on the accuracy of the forecast.

FIGURE 3

Following Granger and Pesaran (2000), we now estimate the forecast probability $\Pr(y_{n+1} < 0)$. Since the error term is assumed to be normally distributed, we have

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

where $\Phi(\cdot)$ denotes the standard-normal c.d.f.

Let $\boldsymbol{d}(\boldsymbol{\eta})$ denote the bias $\mathrm{E}(\boldsymbol{W}\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})$ and $\boldsymbol{V}(\boldsymbol{\eta})$ the variance of $\boldsymbol{W}\hat{\boldsymbol{\eta}}$. Then,

$$\hat{y}_{n+1} \sim \left(\boldsymbol{x}'_{n+1}\boldsymbol{\beta} + \boldsymbol{z}'_{n+1}\boldsymbol{\gamma} - \sigma\boldsymbol{\omega}'\boldsymbol{d}(\boldsymbol{\eta}),\ \sigma^2\left(\boldsymbol{x}'_{n+1}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_{n+1} + \boldsymbol{\omega}'\boldsymbol{V}(\boldsymbol{\eta})\boldsymbol{\omega}\right)\right)$$

so that an approximate 95% confidence interval for the parameter $(\boldsymbol{x}'_{n+1}\boldsymbol{\beta} + \boldsymbol{z}'_{n+1}\boldsymbol{\gamma})/\sigma$ is given by

$$\frac{\hat{y}_{n+1}}{\sigma} + \boldsymbol{\omega}'\boldsymbol{d}(\boldsymbol{\eta}) \pm 1.96\sqrt{\boldsymbol{x}'_{n+1}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_{n+1} + \boldsymbol{\omega}'\boldsymbol{V}(\boldsymbol{\eta})\boldsymbol{\omega}}.$$

This interval still depends on σ and $\boldsymbol{\eta}$ which are unknown. We obtain an estimated confidence interval by replacing them by the estimates $\hat{\sigma}$ and $\hat{\boldsymbol{\eta}}$.

The estimated probability $\Pr(y_{n+1} < 0)$ is given by the solid line in Figure 3. The two dotted lines represent the standard 95% confidence interval for these forecast probabilities. The two dashed lines represent the estimated 95% confidence interval for the forecast probabilities, where pretesting is taken into account.

While the standard regression confidence intervals are already large, allowing only three years (1992, 2000, 2001) where a direction can be forecasted with any confidence, the (correct) pretest confidence intervals are such that we cannot be confident in *any* year. This confirms our general conclusions from Figures 1 and 2, that the accuracy of the forecast is significantly less than usually reported, as a consequence of the fact that pretesting is not usually properly accounted for.

We conclude that ignoring the effects of pretesting on the distribution of the forecast can lead to a very 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 estimator

We have seen that in evaluating the properties of forecasts, 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 solution to this problem is proposed (in the estimation context) for the case $m = 1$. In that case there are only two possible models, the restricted (r) and the unrestricted (u). The forecast then takes the simple form (see (2))

$$\hat{y}_{n+1} = \lambda \hat{y}^{(u)}_{n+1} + (1-\lambda)\hat{y}^{(r)}_{n+1}, \tag{5}$$

where the weight-function $\lambda = \lambda(\hat{\eta})$ is defined as

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

and the prior π as the 'neutral' Laplace density,

$$\pi(\eta) = \frac{c}{2}\exp\left(-c\,|\eta|\right), \quad -\infty < \eta < \infty, \quad c = \log 2. \tag{6}$$

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.

If $m > 1$, then it is not so clear how the weights should be taken. However, in the special case where $\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z} = \boldsymbol{I}_m$, the multi-dimensional problem separates into m one-dimensional problems, and we use the Laplace weights for each dimension separately.

MORE TO FOLLOW HERE

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

6 The case of unknown σ^2

So far we have assumed that σ^2 is known. This is of course unrealistic and we have to show how to obtain the pretest forecast properties also when σ^2 is not known. We shall deal with two cases. First the case where σ^2 is estimated from the residuals in the fully unrestricted model, irrespective from which model is finally selected. This procedure makes the analysis a little easier. Next the more usual case where σ^2 is estimated from the residuals in the selected model.

There is no theoretical problem in doing the calculations, because in either case the estimator $\hat{\sigma}^2$ will depend on $\boldsymbol{M}\boldsymbol{y}$, so that Theorem 1 still applies.

The conclusion is that the qualitative results are unchanged.

7 Concluding remarks

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.

Data appendix

To follow

References

Abadir, K.M. and J.R. Magnus (2002). Notation in econometrics: a proposal for a standard. *The Econometrics Journal* 5, 000–000.

Danilov, D.L. and J.R. Magnus (2001). On the harm that pretesting does. Submitted for publication.

Granger, C.W.J. and M.H. Pesaran (2000). Economic and statistical measures of forecast accuracy. *Journal of Forecasting* 19, 537–560.

Magnus, J.R. (1999). The traditional pretest estimator. *Theory of Probability and Its Applications* 44, 293–308.

Magnus, J.R. (2002). Estimation of the mean of a univariate normal distribution with known variance. *The Econometrics Journal* 05, 000–000.

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.

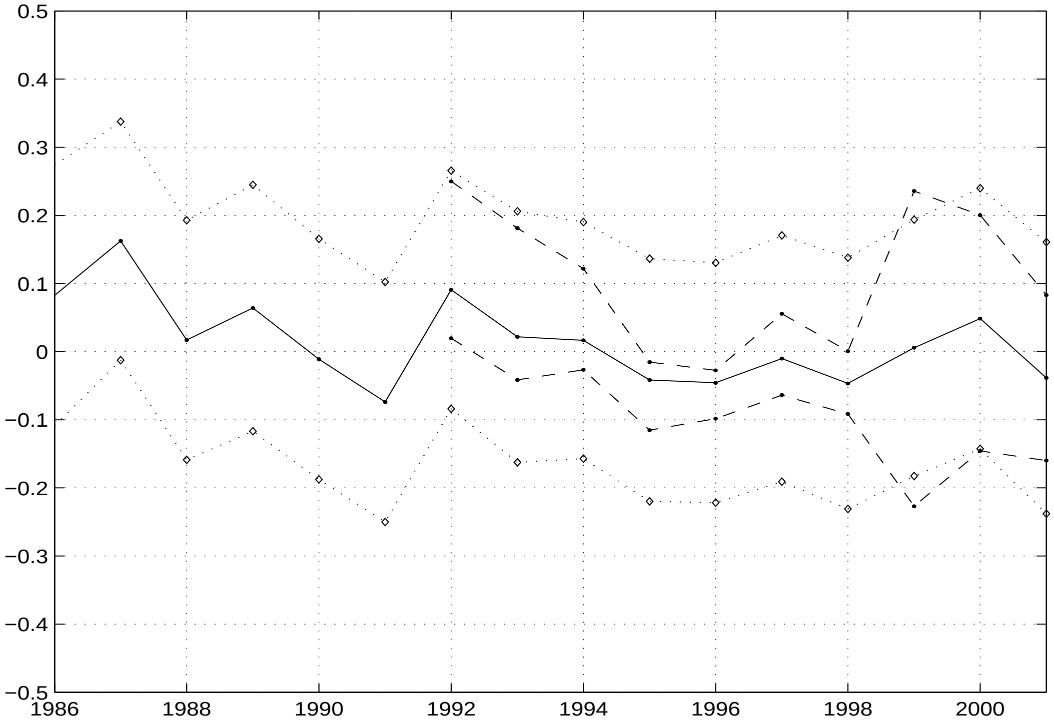

Figure 1. Bias of forecast as a result of pretesting.

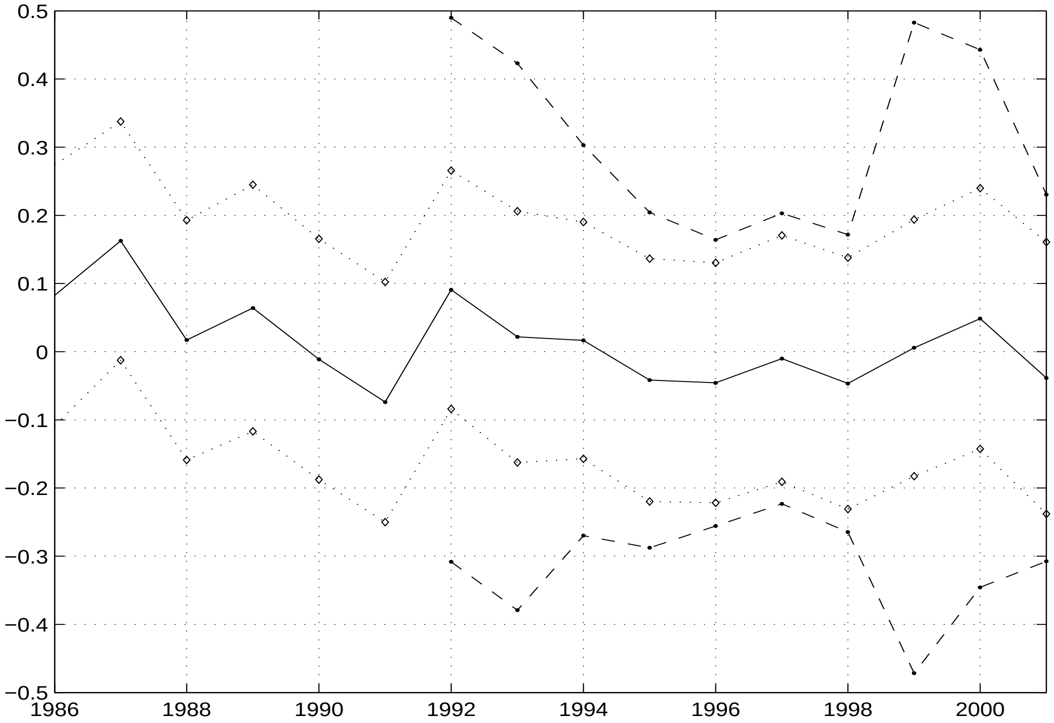

Figure 2. Standard deviation of forecast as a result of pretesting.

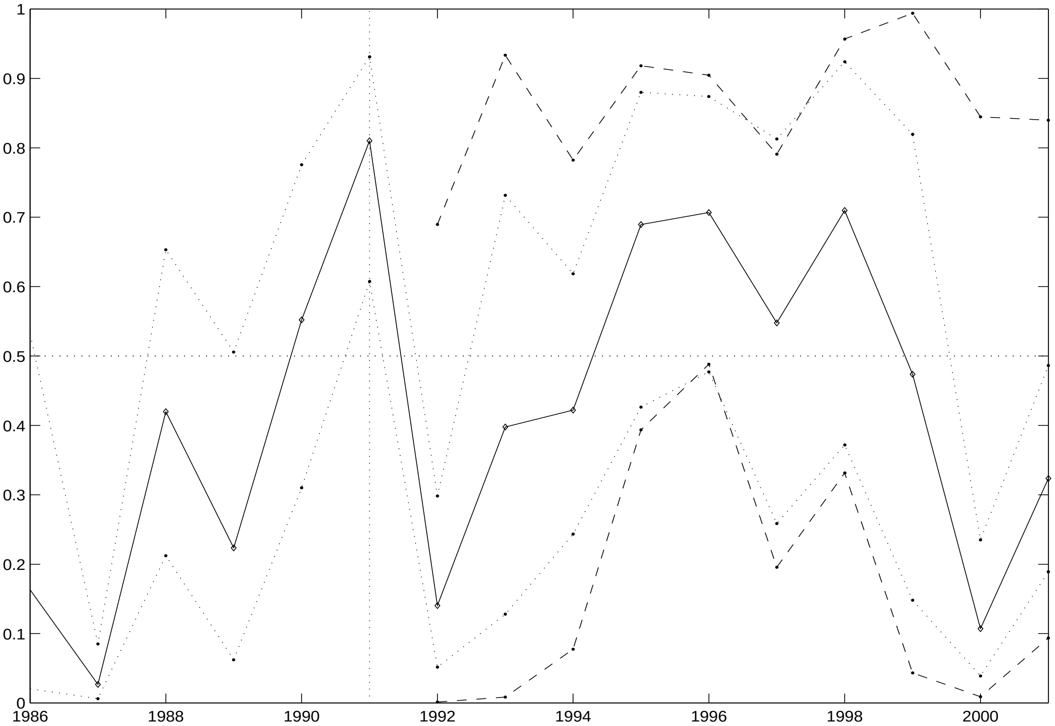

Figure 3. Estimated forecast probabilities.