A NOTE ON USING CROSS-SECTIONAL INFORMATION IN BAYESIAN ESTIMATION OF SECURITY BETAS

OLDRICH A. VASICEK*

BAYESIAN DECISION THEORY provides formal procedures which utilize information available prior to sampling, together with the sample information, to construct estimates which are optimal with respect to the minimization of the expected loss. This paper presents a method for generating Bayesian estimates of the regression coefficient of rates of return of a security against those of a market index. The distribution of the regression coefficients across securities is used as the prior distribution in the analysis. Explicit formulas are given for the estimates. The Bayesian approach is discussed in comparison with the current practice of sampling-theory procedures.

I. INTRODUCTION

The Capital Asset Pricing Model of Treynor [7], Sharpe [6], and Lintner [4] states that the expected rate of return on a security in excess of the risk-free rate is proportional to the slope coefficient of the regression of that security's rates of return on a market index. The slope coefficient, or beta, is for this reason one of the basic concepts of modern capital market theory, and considerable attention has been devoted to its measurement.

Customarily, beta is estimated from past data by least-squares regression procedures. The least-squares technique consists of fitting a linear relationship between the rates of return on a security and the rates of return on a market index so that the sum of squared differences between the security's actual returns and those implied by the relationship is minimized.

If y_t, $t = 1,2, \ldots, T$ and x_t, $t = 1,2, \ldots, T$ are the series of rates of return on a security and on a market index, respectively, the least-squares estimates of the parameters β, α, σ^2 in the simple linear regression process

$$y_t = \alpha + \beta x_t + e_t, \quad t = 1,2,\ldots,T \tag{1}$$

$$Ee_t = 0, \quad Ee_te_s = 0 \quad \text{for} \quad t \neq s, \quad Ee_t^2 = \sigma^2$$

are given as

$$b = \Sigma(y_t - \bar{y})(x_t - \bar{x})/\Sigma(x_t - \bar{x})^2 \tag{2}$$

$$a = \bar{y} - b\bar{x} \tag{3}$$

$$s^2 = \frac{1}{T-2}\Sigma(y_t - a - bx_t)^2, \tag{4}$$

respectively, and the variance of b is estimated as

* Wells Fargo Bank, N.A. This paper is a minor revision of the author's unpublished memorandum "Bayesian Estimates of Beta," Wells Fargo Bank, August 1971.

$$s_b^2 = s^2/\Sigma(x_t - \bar{x})^2. \tag{5}$$

These are the best unbiased estimates of the parameters in the sense that the expected value of each of the estimates is equal to the corresponding parameter and the expected quadratic error attains the minimal value. In particular, when the beta coefficient of a stock is estimated by b, the following holds:

$$E(b|\beta) = \beta. \tag{6}$$

$$\text{Var}(b|\beta) = \text{minimum over all estimates of } \beta \text{ satisfying (6).} \tag{7}$$

For these reasons, the sampling-theory estimation procedures are commonly applied to the estimation of the beta of a security. Yet, the criteria as represented by Equations (6) and (7) do not satisfactorily reflect the desired properties of a beta estimator. Equation (6) describes an aspect of the distribution of the estimate assuming that the true value of the parameter is given. The actual situation is just the reverse: it is the sample coefficient that is known, and on the basis of this (and any prior or additional) information we want to infer about the distribution of the parameter.

To illustrate this point, assume that the estimated beta of a stock traded on the New York Stock Exchange is b = .2. In the absence of any additional information, this value is taken by sampling theory as being the best estimate of the true beta *because* any given true beta is equally likely to be overestimated as underestimated by the sample b. This, however, does not imply that given the sample estimate b, the true parameter is equally likely to be below or above the value .2. In fact, it is known from previous measurements that betas of stocks traded on the New York Stock Exchange are concentrated around unity, and most of them range in value between .5 and 1.5. Thus, an observed beta as low as 0.2 is more likely to be a result of underestimation than overestimation. The question of whether the estimate b is equally likely to lie below or above the true beta is irrelevant, since the true beta is not known. What is desired is an estimate such that given the sample information (which is available), the true beta will with equal probability lie below or above it.

To pursue this example further, assume that there are 1000 stocks under consideration, the betas of which are known to be distributed approximately normally around 1.0 with standard deviation of .5. Each of these true betas is equally likely to be underestimated or overestimated by b. Therefore, there are 500 stocks with true beta higher than the observed estimate, and 500 with true beta lower than the estimate. If an estimate of b = .2 is observed, the stock might be any of the approximately 500 × .945 = 473 stocks with β larger than .2 and underestimated, or any of the approximately 500 × .055 = 27 stocks with β smaller than .2 and overestimated. Apparently, given the sample and our prior knowledge of beta distribution, the former is much more likely, and thus, it is not correct to take .2 for an unbiased estimate.

This has been recognized before in the special situation where portfolios were formed by ranking of sample estimates (cf. Wagner and Vasicek [8]). The knowledge of the cross-sectional distribution of betas, however, can be

used as a prior information whenever a beta of a security is estimated. Also, as a referee pointed out to the author, a similar problem has been recently addressed by Bogue [1]. Following is a Bayesian analysis of the simple normal regression process with the cross-sectional prior information. For information about the principles and techniques of Bayesian statistical theory, the reader is referred to Raiffa and Schlaifer [5].

II. Bayesian Estimates

For computational convenience, reparametrize the regression process (1) as follows:

$$y_t = \eta + \beta(x_t - \bar{x}) + e_t, \quad t = 1,2,\ldots,T \tag{8}$$

where

$$\eta = \alpha + \beta\bar{x}.$$

Assuming normal distribution of the disturbances, the kernel $k(b,\bar{y},s|v,\beta,\eta,\sigma)$ of the likelihood is proportional to (see [5], p. 335)

$$\sigma^{-T}\exp[-(T-2)s^2/(2\sigma^2)].\exp\left[-\frac{1}{2\sigma^2}(T(\bar{y}-\eta)^2 + v(b-\beta)^2)\right] \tag{9}$$

where b, s^2 is given by Equations (2), (4),

$$\bar{y} = \frac{1}{T}\Sigma y_t,$$

and

$$v = \Sigma(x_t - \bar{x})^2.$$

Let the information available prior to sampling consist of knowledge of the cross-sectional distribution of betas. Assuming that the distribution is approximately normal with parameters b′, s'_b, the marginal prior density of β is

$$f'(\beta) \propto \exp[-(\beta - b')^2/(2s'_b{}^2)]. \tag{10}$$

(In accordance with practice, the prior distributions and parameters are denoted by primed letters, the posterior by letters with double primes, and the sample information without superscripts.)

Unless some prior information is available on η, σ, it is assumed that the prior density of these parameters is assessed as

$$f'(\eta,\sigma) \propto \sigma^{-1} \tag{11}$$

and independent of $f'(\beta)$. The density (11) is an improper density function corresponding to the limiting case where the prior information on η, σ is totally negligible. The joint prior density of the parameters β, η, σ is then

$$f'(\beta,\eta,\sigma) \propto \sigma^{-1}.\exp[-(\beta - b')^2/(2s'_b{}^2)]. \tag{12}$$

Note that the prior distribution (12) is not of the natural conjugate form (the bivariate normal-gamma distribution for the simple normal regression

process). The reason why the natural conjugate density is not suitable here is that the conjugate prior expresses prior information in the form as if it were results of previous sampling from the same process, and it is not rich enough to give a good representation of the case when the prior information involves a cross-sectional relationship among several regression processes.

Given the prior density (12), the posterior density f″ of the parameters β, η, σ is evaluated using Bayes' theorem:

$$f''(\beta, \eta, \sigma|v, b, \bar{y}, s) \tag{13}$$
$$= f'(\beta, \eta, \sigma)k(b, \bar{y}, s|v, \beta, \eta, \sigma)N(b, \bar{y}, s)$$

where

$$N^{-1}(b, \bar{y}, s) = \int f'(\beta, \eta, \sigma)k(b, \bar{y}, s|v, \beta, \eta, \sigma)d\beta\, d\eta\, d\sigma.$$

The marginal posterior density of β is evaluated as

$$f''(\beta|v, b, \bar{y}, s) = \int f''(\beta, \eta, \sigma|v, b, \bar{y}, s)d\eta\, d\sigma.$$

After substitution, this yields

$$f''(\beta|v, b, \bar{y}, s) \propto \exp[-(\beta - b')^2/(2s'_b{}^2)]. \left[T - 2 + \frac{v(\beta - b)^2}{s^2}\right]^{-\frac{1}{2}(T-1)}. \tag{14}$$

When T is larger than 20, the posterior distribution of β is approximately normal with mean b″ and variance $s''_b{}^2$, where

$$b'' = \frac{b'/s'_b{}^2 + b/s_b{}^2}{1/s'_b{}^2 + 1/s_b{}^2} \tag{15}$$

$$s''_b{}^2 = \frac{1}{1/s'_b{}^2 + 1/s_b{}^2}. \tag{16}$$

Here

$$s_b{}^2 = s^2/v$$

is the estimated variance of b as given by Equation (5). (In sampling-theory terminology, s_b is usually called the standard error of the estimate b.)

The marginal posterior density of β describes the knowledge about the distribution of the estimated parameter, given the information from the sample and the prior information. The choice of a point estimate of β depends on this posterior distribution as well as the utility function on the space of decisions (estimates). Under a quadratic terminal loss function (which is a Bayesian analogue to the sampling-theory concept of minimum variance estimates) the optimal estimate of β is the mean of the posterior distribution (14). For $T > 20$, the error of approximating the posterior mean by b″ does not exceed .01 and decreases approximately linearly with 1/T. Since this error is small in comparison with the dispersion s''_b of the posterior distribution, no material

loss is incurred when b'' is taken for the estimate that minimizes the expected quadratic opportunity loss.

III. Discussion and Conclusions

The Bayesian estimate b'' as given by Equation (15) can be interpreted as an adjustment of the sample estimate b toward the best prior estimate b', the degree of adjustment being proportionate to the precision $h = 1/s_b^2$, $h' = 1/s'^2_b$ of the sample estimate and the prior distribution, respectively. Equation (16) can be interpreted as stating that the precision $h'' = 1/s''^2_b$ of the posterior distribution is the sum of the precision of b and that of the prior distribution.

The choice of the parameters b', s'_b of the prior density $f'(\beta)$ depends on the prior information available. If nothing is known about a stock prior to sampling except that it comes from a certain population of stocks (for instance, from the population of all stocks traded on the New York Stock Exchange), an appropriate choice of the prior density is the cross-sectional distribution of betas observed for that population. For the New York Stock Exchange population, the prior parameters might be approximately $b' = 1$, $s'_b = .5$. In this case, the regression coefficient estimated from the sample is linearly adjusted toward unity, the degree of the adjustment depending upon the standard error s_b of the estimate.

A somewhat similar procedure is used in the Security Risk Evaluation service by Merrill Lynch, Pierce, Fenner & Smith, Inc. Their simplified method utilizes a formula of the form

$$b'' = 1 + k(b - 1) \tag{17}$$

where k is a constant common for all stocks. This constant can be interpreted as the slope of the cross-sectional regression of beta estimates on those obtained over a prior non-overlapping period. Comparison of Equation (17) with Equation (15) shows that this method assumes that the variance s_b^2 of the sample regression coefficient is the same for all securities. The effect of this procedure is thus to overadjust more accurate estimates and underadjust the less accurate ones.

In some cases, more can be known about a stock than that it comes from a certain population. Assume, for instance, that a stock is selected on the basis of an instrumental variable, which may be related to the true betas but not to the estimation error of the sample estimates b. In this case, a proper choice of the prior distribution is the distribution of betas implied by the knowledge of the instrumental variable. Thus, if a utility stock is considered, and t is known from previous measurements that betas of utilities are centered around .8 with a dispersion of .3, the estimate b is adjusted toward .8 by the formula (15) with $b' = .8$, $s'_b = .3$. In general, the degree and direction of the adjustment depend on the prior distribution $f'(\beta)$ as characterizing the information pertaining to β that is contained in the instrumental variable.

When estimating beta of a portfolio composed of N stocks, the sample estimate b is again adjusted through the formula (15). In this case, however, the value used for s'_b is the cross-sectional dispersion of betas of portfolios of size N.

In most instances, a good approximation for this dispersion is obtained by assuming cross-sectional independence of the regression residuals (as in the diagonal model) and consequently using the cross-sectional dispersion of individual securities' betas reduces by the factor of $1/\sqrt{N}$.

In some cases, the prior information may contain information of another sample from the same process (as, regression results over a previous period) but the two samples cannot be pooled. This situation arises, for example, when a portfolio is formed by ranking securities on the basis of their estimated betas and then the portfolio's beta is estimated over the next period. In such cases, the estimation proceeds in two steps. First, the posterior distribution based on the first sample and the cross-sectional prior is obtained. Next, this posterior distribution is used as the prior density to utilize the information of the second sample. Thus, the sample estimate from the second sample is adjusted toward the adjusted first sample estimate.

In summary, the estimate of a security's beta that minimizes the expected squared estimation error is given by Equation (15), where the parameters b', s'_b of the prior distribution are chosen to reflect *all* the information on beta available prior to sampling. The mean squared estimation error s''^2_b is given by Equation (16).

The relative merit of this Bayesian estimation method as contrasted to procedures of sampling theory will now be briefly discussed. The main objection to the Bayesian estimation method is that the estimate b'' is not an unbiased estimate of β (in the sampling-theory sense), while b is unbiased,

$$E(b''|\beta) \neq \beta,$$

$$E(b|\beta) = \beta. \tag{18}$$

To discuss this objection, it is useful to ask why unbiasedness in the sense of Equation (18) is desirable. One can identify two reasons, the first of which is that, in virtue of the law of large numbers, an unbiased estimate converges in probability to the estimated parameter as the sample size increases,

$$\operatorname*{Plim}_{T\to\infty} b = \beta.$$

The same, however, is true for the estimate b'',

$$\operatorname*{Plim}_{T\to\infty} b'' = \beta,$$

since with increasing sample size $s_b^2 \to 0$ and the degree of the adjustment decreases.

The second reason for requiring an unbiased estimate is that the mean quadratic error

$$E((\hat{\beta} - \beta)^2|\beta) \tag{19}$$

is minimized in a class of estimates $\hat{\beta}$ of the same variance by an unbiased estimate. The expected value (19) is taken with respect to the conditional likelihood (9) of the sample. This, however, is not justified. Rather than minimizing the squared *sampling* error, what should be done is to minimize the squared *estimation* error. That is, minimize

$$E''(\hat{\beta} - \beta)^2, \tag{20}$$

the expectation being taken with respect to the posterior distribution of β. The estimate b″, not b, is the estimate $\hat{\beta}$ to minimize (20).

This is more than a mere philosophical point. If two persons, one using the estimate b and the other b″, were penalized proportionally to the squared difference of their respective estimates from the true parameter value β (or, for that matter, from the next-period sample estimate), the former would go broke first.

In conclusion, Bayesian estimates (15) are preferred to the classical sampling-theory estimates (2) for the following reasons: First, Bayesian procedures provide estimates that minimize the loss due to misestimation, while sampling-theory estimates minimize the error of sampling. This is because Bayesian theory deals with the distribution of the parameters given the available information, while sampling theory deals with the properties of sample statistics given the true value of the parameters. Secondly, Bayesian theory weights the expected losses by a prior distribution of the parameters, thus incorporating knowledge which is available in addition to the sample information. This is particularly important in the case of estimating betas of stocks, where the prior information is usually sizeable.

REFERENCES

1. Marcus C. Bogue. "The Estimation and Behavior of Systematic Risk," unpublished dissertation, Graduate School of Business Administration, Stanford University (1972).
2. Basil A. Kalymon. "Estimation Risk in the Portfolio Selection Model," *Journal of Financial and Quantitative Analysis* (January 1971).
3. Michael Kantor. "Market Sensitivities," *Financial Analysts Journal* (January 1971).
4. John Lintner. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics* (February 1965).
5. Howard Raiffa and Robert Schlaifer. *Applied Statistical Decision Theory* (Harvard University, Boston, 1961).
6. William F. Sharpe. "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance* (September 1964).
7. Jack L. Treynor. "Toward A Theory of Market Value of Risky Assets," unpublished memorandum (1961).
8. Wayne H. Wagner and Oldrich A. Vasicek. "The Effect of Estimation Error of Beta on the Risk of Passive Portfolios," unpublished memorandum, Wells Fargo Bank (March 1971).