The Effect of Model Misspecification on Tests of the Efficient Market Hypothesis

Menachem Brenner

Journal of Finance, Volume 32, Issue 1 (Mar., 1977), 57-66.

THE JOURNAL OF FINANCE · VOL. XXXII, NO. 1 · MARCH 1977

THE EFFECT OF MODEL MISSPECIFICATION ON TESTS OF THE EFFICIENT MARKET HYPOTHESIS

MENACHEM BRENNER*

I. INTRODUCTION

TESTS OF THE Efficient Market Hypothesis (EMH) are in general "weak" tests. The null hypothesis has always been that the market is efficient with no specific alternative of inefficiency. Thus, the power of these tests is not known. The tests usually rely on a certain market model without questioning the validity of the model that was used. A misspecified model may provide test statistics that indicate that the market is efficient when it is not efficient and vice versa. The possibility that the EMH has not been rejected because the wrong market model was used, was never adequately considered.

This paper is concerned with the theoretical effect of model misspecification on tests of the efficient market hypothesis. Since tests of the EMH usually use a certain market model, the conclusions are based on the assumption that the model is correctly specified. By deriving the possible biases due to model misspecification, we are actually concerned with the power (or validity) of the EMH tests.

Tests of the EMH generally proceed in two stages: First, we estimate the relevant parameters using a certain market model; second, we use the estimated parameters for prediction and use the prediction errors, also called "residuals",[1] to test market efficiency.

The statistical properties of the parameters, estimated in the first stage, depend on how well the market model describes the true underlying stochastic process. Serious misspecifications may yield biased and/or inefficient parameter estimates. This in turn may result in biased and/or inefficient estimates of the residuals in the second stage. The extent to which these misspecifications affect our conclusion about market efficiency depends on the way we use these residuals in testing the EMH. Under certain circumstances (to be specified later) the misspecifications, no matter how serious, will not affect our conclusions with regard to market efficiency.

In the remainder of this paper we first consider general cases of misspecification, their effect on parameter estimates and on residual estimates. Then we consider the effects of misspecification in some specific cases. Finally we present the effect of β changes on residual estimates.

* Hebrew University, Jerusalem. This paper is extracted from my Ph.D. thesis at Cornell University. I wish to express my indebtness to my thesis committee, Seymour Smidt, Bernell Stone, and John McClain. Special thanks are due to Professor Seymour Smidt for his help and insightful comments. This research has been partially supported by a grant from the Lady Davis Foundation.

1. Those are not the residuals of the regression equation obtained in the first stage.

II. Model Misspecification: General[2,3]

Since research on efficient markets centers around one or two factor market models we will show the biases (or the lack of biases) from using the one-factor model when a two-factor model is a correct specification and vice versa. The well-known one-factor market model is given by:

$$\tilde{R}_{jt} = \alpha_j + \beta_j \tilde{R}_{mt} + \tilde{u}_{jt} \tag{1}$$

where $\tilde{u}_{jt}$ satisfies all the necessary assumptions (if (1) is the correct specification). A two-factor market model can be stated as

$$\tilde{R}_{jt} = a_j + b_j \tilde{R}_{mt} + c_j \tilde{R}_{kt} + \tilde{e}_{jt} \tag{2}$$

where a_j, b_j, c_j are constants pertinent to j, $\tilde{R}_{kt}$ represents a second market factor and $\tilde{e}_{jt}$ satisfies all the necessary statistical assumptions (if (2) is the correct specification). If we use the incorrect model to estimate the parameters, we commit a specification error. Using (2) when (1) is correct amounts to including an irrelevant variable while using (1) when (2) is correct amounts to omitting a relevant variable.

A. *Omitting a Relevant Variable*

We first assume that (2) is the correct specification but (1) is used in estimation and prediction. It is well known that a misspecification in the form of an omitted variable yields biases that depend on the correlation between the omitted variable and the existing variables. If $\text{cov}(\tilde{R}_k, \tilde{R}_m) \neq 0$ then $\text{cov}(\tilde{R}_m, \tilde{u}_j) \neq 0$. This will result in biased and inconsistent parameter estimates that may bias the residuals used to test the EMH.[4]

For a specific set of observations T, the estimates (denoted by hats) are

$$\hat{\beta}_j = \frac{\hat{\text{cov}}(\tilde{R}_j, \tilde{R}_m)}{\hat{\sigma}^2(\tilde{R}_m)} = b_j + c_j \frac{\hat{\text{cov}}(\tilde{R}_k, \tilde{R}_m)}{\hat{\sigma}^2(\tilde{R}_m)} + \frac{\hat{\text{cov}}(\tilde{R}_m, \tilde{e}_j)}{\hat{\sigma}^2(\tilde{R}_m)} \tag{3a}$$

$$\hat{\alpha}_j = \overline{R}_j - \hat{\beta}_j \overline{R}_m = a_j + c_j \left[\overline{R}_k - \overline{R}_m \frac{\hat{\text{cov}}(\tilde{R}_k, \tilde{R}_m)}{\hat{\sigma}^2(\tilde{R}_m)} \right] - \tilde{R}_m \frac{\hat{\text{cov}}(\tilde{R}_m, \tilde{e}_j)}{\hat{\sigma}^2(\tilde{R}_m)} \tag{3b}$$

where bars denote sample means. To simplify notation we define

$$\lambda_k \equiv \frac{\text{cov}(\tilde{R}_k, \tilde{R}_m)}{\sigma^2(\tilde{R}_m)} \quad \text{and} \quad \delta_j \equiv \frac{\text{cov}(\tilde{R}_m, \tilde{e}_j)}{\sigma^2(\tilde{R}_m)}$$

2. For clarity of presentation, the analysis of this section is limited to two-variable models.

3. Most econometric issues and their effect on tests of the Capital Asset Pricing Model (CAPM) are considered in Miller and Scholes [8]. A related treatment, but in somewhat different context, is Roll's [9].

4. Tests of the EMH are not necessarily affected. The effect will depend on the test statistic and on some properties of the event studied. These conditions are described later on.

If we use probability limits to get at the bias in $\hat{\beta}$ and $\hat{\alpha}$ we get

$$\operatorname*{plim}_{T\to\infty} \hat{\beta}_j = b_j + c_j\lambda_k \tag{4a}$$

$$\operatorname*{plim}_{T\to\infty} \hat{\alpha}_j = a_j + c_j\left[E(\tilde{R}_k) - E(\tilde{R}_m)\lambda_k\right] \tag{4b}$$

For given values of c_j, $E(\tilde{R}_k)$ and $E(\tilde{R}_m)$, λ_k will determine the bias in $\hat{\beta}_j$ and $\hat{\alpha}_j$. If $\lambda_k = 0$ then, $\hat{\beta}_j$ is a consistent estimate of b_j, and $\hat{\alpha}_j$ is a consistent estimate of $a_j + c_jE(\tilde{R}_k)$.

Since (3a) and (3b) is more general than (4a) and (4b) and, for empirical tests, more accurate we continue with the analysis of (3a) and (3b).[5]

While the estimate $\hat{\lambda}_k$ is common to all companies with a common observation period, the estimate $\hat{\delta}_j$ is pertinent to company *j*. A cross-sectional combination of $\hat{\alpha}$ and $\hat{\beta}$ (like in tests of the EMH) will therefore drive $\hat{\delta}_j$ to zero but will not effect $\hat{\lambda}_k$, an estimate that is based on market factors in the period that the events occur. Since all our tests involve combinations of securities we drop $\hat{\delta}_j$ and rewrite (3a) and (3b) as

$$\hat{\beta}_j = b_j + c_j\hat{\lambda}_k \tag{5a}$$

$$\hat{\alpha}_j = a_j + c_j\left[\overline{R}_k - \overline{R}_m\hat{\lambda}_k\right] \tag{5b}$$

The biases in parameter estimates may affect the residual estimates, obtained in the next stage, thereby affecting tests of the EMH. Again, given the correct specification (2) we continue to use (1) to compute the residuals in the residual period. The estimates $\hat{\alpha}_j$ and $\hat{\beta}_j$ are used to predict R_{jt+i} *conditional* on R_{mt+i} (i denotes the months in the residual period). The residuals are computed by

$$\hat{u}_{jt+i} = R_{jt+i} - \hat{R}_{jt+i} = a_j + b_jR_{mt+i} + c_jR_{kt+i} + e_{jt+i} - \left(\hat{\alpha}_j + \hat{\beta}_jR_{mt+i}\right) \tag{6a}$$

Substituting the Right-Hand-Side (RHS) of (5a) and of (5b) for $\hat{\alpha}_j$ and $\hat{\beta}_j$ in (6a) we get

$$\begin{aligned}\hat{u}_{jt+i} &= a_j + b_jR_{mt+i} + c_jR_{kt+i} + e_{jt+i} - \left(a_j + c_j\left[\overline{R}_k - \overline{R}_m\hat{\lambda}_k\right] + \left[b_j + c_j\hat{\lambda}_k\right]R_{mt+i}\right)\\ &= e_{jt+i} + c_j\left[\left(R_{kt+i} - \overline{R}_k\right) - \hat{\lambda}_k\left(\overline{R}_{mt+i} - \overline{R}_m\right)\right]\end{aligned} \tag{6b}$$

The computed residual $\hat{u}_{jt+i}$ will deviate from the true residual e_{jt+i}, for given values of c_j and $\hat{\lambda}_k$, by the deviations of R_{kt+i} and R_{mt+i} from their respective sample means based on the estimation period.

To test the EMH we combine the residuals in month *i* across all securities in the

5. Since tests of market efficiency center around events that may concentrate at certain time periods (e.g., dividend announcements) and since company parameters are estimated around these events to minimize the negative effects of possible non stationarity in the parameters, it is possible that $\hat{\lambda}_k$ is not even close to λ_k (even for large *T*). Thus, a misspecification may cause $\hat{\beta}_j$ to deviate largely from b_j even if $\lambda_k = 0$.

sample, where i is generally not the same calendar month. The test statistics, most commonly used is

$$\mathrm{AR}_{t+i} \equiv \bar{\hat{u}}_{t+i} = \frac{1}{n} \sum_{j=1}^{n} \hat{u}_{jt+i} \tag{7a}$$

the cross-sectional mean of the residuals at month i after the announcement of new information. The behavior of the cumulative residual

$$\mathrm{CAR}_L \equiv \bar{U}_L = \sum_{i=1}^{L} \bar{\hat{u}}_{t+i} \tag{7b}$$

is then examined.

Using (6b) to compute the average residual AR_{t+i} we get

$$\mathrm{AR}_{t+i} = \frac{1}{n} \sum_{j=1}^{n} \hat{u}_{jt+i} = \frac{1}{n} \sum_{j=1}^{n} \left(e_{jt+i} + c_j \left[\left(R_{kt+i} - \bar{R}_k \right) - \hat{\lambda}_k \left(R_{mt+i} - \bar{R}_m \right) \right] \right.$$

$$= e_{t+i} + \bar{c} \left[\left(\bar{R}_{kt+i} - \bar{\bar{R}}_k \right) - \bar{\hat{\lambda}}_k \left(\bar{R}_{mt+i} - \bar{\bar{R}}_m \right) \right] + \left[\hat{\mathrm{cov}}\left(c_j, R_{kt+i} - \bar{R}_k \right) \right.$$

$$\left. - \hat{\lambda}_k \, \hat{\mathrm{cov}}\left(c_j, R_{mt+i} - \bar{R}_m \right) \right]$$

where the estimates $\bar{\bar{R}}_k, \bar{\bar{R}}_m$ and $\bar{\hat{\lambda}}_k$ are based on estimation periods for all securities in the sample.[6]

The average residual AR_{t+i} in (7b) contains three components. The first component, the cross sectional average residual in month i based on the correctly specified model, is the required variable in tests of the EMH. The other two components are bias terms that may disappear under the following circumstances:

1. If c_j is not correlated with the deviations of R_{kt+i} and R_{mt+i} from their respective means then the last term in (7b) vanishes (this is a sufficient but not necessary condition). In other words, if we have no reason to believe that, for example, companies with high c values participate in the event in periods of large $R_{kt+i} - \bar{R}_k$ (or large $R_{mt+i} - \bar{R}_m$) while companies with low c values do it in periods of small $R_{kt+i} - \bar{R}_k$ (or small $R_{mt+i} - \bar{R}_m$) then we may be willing to assume that the last term is zero.

2. The first bias term (the second term in (7b) will disappear if $\bar{c} = 0$ or if the market factors averages in the prediction period are not significantly different from their means in the estimation period (again this later condition is sufficient but not necessary). If the event studied is fairly well distributed over time and the estimation period is some period prior to or around the prediction period then $\bar{R}_{kt+i} \cong \bar{\bar{R}}_k$ and $\bar{R}_{mt+i} \cong \bar{\bar{R}}_m$. If the event is clustered in time then there is a higher chance that $\bar{R}_{kt+i} \neq \bar{\bar{R}}_k$ and/or $\bar{R}_{mt+i} \neq \bar{\bar{R}}_m$ and the second term will not vanish

6. $\bar{R}_{mt+i}$ is computed as $(1/n)\sum_{j=1}^{n} R_{mt+i}$ where month i refers to the same month relative to the event, but for different securities it may be a different calendar month.

unless $\bar{c}=0$. If the conditions for the disappearance of the bias terms are fulfilled then

$$\mathrm{AR}_{t+i} \equiv \bar{\hat{u}}_{t+i} \cong \bar{e}_{t+i} \tag{7d}$$

Thus, despite the fact that we used the incorrectly specified model we ended up using the correct variable in tests of the EMH. The main reason is of course the offsetting biases in $\hat{\alpha}$ and $\hat{\beta}$. Thus, to get at (7d), in addition to the conditions regarding the bias terms, we also require the incorrect model to be used consistently in the estimation period and in the prediction period.[7] If the market model is not used *consistently* or some of the conditions for zero bias are not fulfilled, we will have

$$B(u,e)_i \equiv \bar{\hat{u}}_{t+i} - \bar{e}_{t+i} \neq 0 \tag{7e}$$

If the sample consists of many stocks over a long period of time then we expect that $B(u,e)_i$ will be of roughly the same magnitude at all months i (the prediction period). Therefore, the cumulative bias

$$\mathrm{CB}(u,e)_L \equiv \bar{U}_L - \bar{E}_L = \sum_{i=1}^{L} \bar{\hat{u}}_{t+i} - \sum_{i=1}^{L} \bar{e}_{t+i} \tag{7f}$$

should increase with L (the number of months used to get (7f)). For example, if $L=24$ we expect that $\mathrm{CB}(u,e)_L \cong B(u,e)_i \cdot 24$. Thus, if $\bar{E}_L$ (the cumulative average residual from the correct model) exhibits a flat line (i.e., an efficient market) $\bar{U}_L$ should be rising or falling, depending on the sign of $B(u,e)_i$, indicating the existence of inefficiency in the market.

B. *Omitting R_{ft} or $\tilde{R}_{zt}$*

Since the previous section is more general, we have not been explicit about the second market factor—R_{kt}. The generality of the model makes the analysis applicable to any specific second factor and to any constraints placed on the coefficients. Since two factor market models that use R_{ft} or $\tilde{R}_{zt}$ as the second factor and the single factor model represent, almost, the entire spectrum of models that are used in EMH studies, we concentrate on the analysis of these specific models. Since the analysis of (2) with either R_{ft} or $\tilde{R}_{zt}$ replacing $\tilde{R}_{kt}$, is very similar we continue to use $\tilde{R}_{kt}$ to represent R_{ft} or $\tilde{R}_{zt}$. For convenience and consistency with other studies we use, from now on, $\tilde{\gamma}_{0t}$ instead of $\tilde{R}_{zt}$.

If R_{kt} in (2) is $\tilde{\gamma}_{0t}$ and a link with a CAPM is not required[8] (i.e., no constraints are placed on the coefficients) then the former analysis retains its generality. However, if we specify (2) in a manner consistent with a CAPM we can restate (2) as

$$\tilde{R}_{jt} = \beta_j \tilde{R}_{mt} + (1-\beta_j)\tilde{R}_{kt} + \tilde{e}_{jt} \tag{8}$$

7. Since using market models inconsistently became a common practice recently, we show later on the danger of such a methodology for tests of the EMH.

8. Kaplan and Roll [6] use (2) with R_{ft} as $\tilde{R}_{kt}$ and place no constraints on the coefficients.

This is simply achieved by placing constraints on the coefficients in (2).[9] If we now assume that (8) is the correct specification but (1) is used we can state the potential biases in $\hat{\alpha}$ and $\hat{\beta}$ by simply using the more general expressions (5a) and (5b) where a_j, b_j, and c_j are replaced by $0, \beta_j$ and $(1-\beta_j)$ respectively. We then have

$$\hat{\beta}_j = \beta_j + (1-\beta_j)\hat{\lambda}_k \tag{9a}$$

$$\hat{\alpha}_j = (1-\beta_j)\left[\bar{R}_k - \bar{R}_m \hat{\lambda}_k\right] \tag{9b}$$

If $\hat{\lambda}_k = 0$, then

$$\hat{\beta}_j = \beta_j \qquad \hat{\alpha}_j = (1-\beta_j)\bar{R}_k$$

and

$$\hat{u}_{jt+i} = (1-\beta_j)\left(R_{kt+i} - \bar{R}_k\right) + e_{jt+i} \tag{9c}$$

where, as before, t refers to the estimation period and i is in the prediction period. (9c) is, of course, conditional on the realized market factors. As far as tests of the EMH are concerned $\bar{\hat{u}}_{t+i} = \bar{e}_{t+i}$ as long as the means of R_k in the prediction and estimation periods are equal or $\bar{\beta} = 1$ (neglecting $\hat{\text{cov}}(\beta_j, R_{kt+i} - \bar{R}_k)$).

If, however, $\hat{\lambda}_k \neq 0$ then the biases in $\hat{\beta}_j$ and $\hat{\alpha}_j$ depend on $\hat{\lambda}_k, \bar{R}_k, \bar{R}_m$ and β_j. If $\hat{\lambda}_k > 0 (<0)$ the bias term in $\hat{\beta}_j$ will be positive (negative) when $\beta_j < 1 (>1)$ and negative(positive) when $\beta_j > 1 (<1)$ but the bias reverses for $\hat{\alpha}_j$ (provided $\bar{R}_m > 0$).[10]

To analyze the effect of the biases in the parameters on tests of the EMH we can refer to corresponding equations in the previous general section. The bias in $\hat{u}_{jt+i}$, given by equation (6b) is restated here as

$$\hat{u}_{jt+i} = e_{jt+i} + (1-\beta_j)\left[\left(R_{kt+i} - \bar{R}_k\right) - \hat{\lambda}_k\left(R_{mt+i} - \bar{R}_m\right)\right] \tag{11a}$$

where c_j is replaced by $1-\beta_j$. The average residual given by equation (7b) is

$$\bar{\hat{u}}_{t+i} = \bar{e}_{t+i} + (1-\bar{\beta})\left[\left(\bar{R}_{kt+i} - \bar{\bar{R}}_k\right) - \bar{\hat{\lambda}}_k\left(\bar{R}_{mt+i} - \bar{\bar{R}}_m\right)\right] \tag{11b}$$

9. Equivalently, we can assume that the following process prevails:

$$\tilde{R}_{jt} = \tilde{\gamma}_{0t} + \tilde{\gamma}_{1t}\beta_{jt} + \tilde{v}_{jt} \tag{8a}$$

Invoking all the assumptions used in the "simple" market model, (8a) is also true for the market portfolio. Averaging (8a) across all j we get

$$\tilde{R}_{mt} = \tilde{\gamma}_{0t} + \tilde{\gamma}_{1t} \tag{8b}$$

Substitute (8b) in (8a) to get

$$\tilde{R}_{jt} = \beta_j \tilde{R}_{mt} + (1-\beta_j)\tilde{\gamma}_{0t} + \tilde{v}_{jt} \tag{8c}$$

When OLS is applied to (8a) to get the estimates $\hat{\gamma}_{0t}$ and $\hat{\gamma}_{1t}$ it must be that $R_{mt} = \hat{\gamma}_{0t} + \hat{\gamma}_{1t}$ because $\sum_j \hat{v}_{jt} = 0$, we can then safely use $R_{mt} - \hat{\gamma}_{0t}$ instead of $\hat{\gamma}_{1t}$.

10. Roll [9] deals with asymptotic biases due to $\text{cov}(\tilde{R}_f, \tilde{R}_m) \neq 0$ and points out the direction of the biases. He does not, however, deal with their possible effects on tests of the EMH.

where $\bar{c}$ is replaced by $1-\beta$ and the last term in (7b) is neglected. $\bar{\beta}=1$ is a sufficient condition for the bias term to vanish.[11] The rest of the analysis follows exactly the analysis in the previous section.[12]

C. *Using Market Models Inconsistently*

Thus far we have assumed that the same model is used in estimation and prediction. We have specified the conditions under which two different models (one correctly specified and the other misspecified) will provide identical average residuals. In our opinion these conditions are very mild and will be mostly met. Thus, if the researcher believes that (8) is generally correct but does not know what variable to use as a second factor he may be better off using the single-factor market model all along. Recently, however, it became a general practice to use one market model for estimation and another for prediction (see Ball [1], Jaffe [5], Mandelker [7]). It is generally assumed that (8) is the governing process and therefore (8) is used for prediction but (1) is used to estimate β. To analyze the effect of such a procedure we also assume (8) to be correct and use (1) to obtain $\hat{\beta}_j$. The residual e_{jt+i} is estimated using (8) and $\hat{\beta}_j$ as follows:

$$\hat{e}_{jt+i}=R_{jt+i}-\hat{R}_{jt+i}=(1-\beta_j)R_{kt+i}+\beta_jR_{mt+i}+e_{jt+i}-\left(1-\hat{\beta}_j\right)R_{kt+i}-\hat{\beta}_jR_{mt+i} \tag{12a}$$

Using (9a) for β_j we have

$$\begin{aligned}\hat{e}_{jt+i}&=(1-\beta_j)R_{kt+i}+\beta_jR_{mt+i}+e_{jt+i}\\&\quad-\left[1-\beta_j-(1-\beta_j)\hat{\lambda}_k\right]R_{kt+i}-\left[\beta_j+(1-\beta_j)\hat{\lambda}_k\right]R_{mt+i}\\&=\hat{\lambda}_k(1-\beta_j)\left[R_{kt+i}-R_{mt+i}\right]+e_{jt+i}\end{aligned} \tag{12b}$$

The average computed residual is then

$$\bar{\hat{e}}_{t+i}=\hat{\lambda}_k(1-\bar{\beta})\left[\bar{R}_{mt+i}-\bar{R}_{kt+i}\right]+\bar{e}_{t+i} \tag{12c}$$

If $\hat{\lambda}_k=0$ or $\bar{\beta}=1$ then $\bar{\hat{e}}_{t+i}$ is unbiased, but in many cases these conditions do not hold. For example, if we use, as before $\hat{\lambda}_f<0$ and $\bar{\beta}<1$ (like in stock splits), then $\bar{\hat{e}}_{t+i}$ is biased upward and the CAR using $\bar{\hat{e}}_{t+i}$ should be increasing when the CAR of the true residual is flat.

The sign and size of $\hat{\lambda}_k$ depend on the study at hand. If we are willing to accept

11. In most tests $\bar{\hat{\beta}}$ was significantly different from 1. (e.g. see Fama, Fisher, Jensen and Roll [3]).

12. In this section we have analyzed the inclusion (or rather exclusion) of R_{ft}, or γ_{0t}, by the same equation (8) since these variables play a similar role in a market model like (8). But there is some difference that might distort the previously obtained results. While R_{ft} is not a random variable and is assumed to be measured without error, $\tilde{\gamma}_{0t}$ is an unobserved random variable, that is replaced by the estimate $\hat{\gamma}_{0t}$. $\hat{\gamma}_{0t}$ was obtained from a regression using (8a) where β_j was estimated using (1) (see Fama-MacBeth [11]). If we assume that $\hat{\gamma}_{0t}$ is an unbiased estimate of $E(\tilde{\gamma}_{0t})$ then the analysis of this section is only affected when $\tilde{\gamma}_{0t}$ is replaced by $\hat{\gamma}_{0t}$ in estimating β_j (due to "errors in variables").

the claim that $\hat{\lambda}_k$ is generally very small in magnitude, especially in studies that spread over a long time period, then this procedure (of combining models) will not affect tests of the EMH. Nevertheless, a somewhat more complicated (but maybe more realistic) situation may affect our tests even when $\hat{\lambda}_k=0$. Assume that a specific version of (8)—($\tilde{\gamma}_{0t}$) is the correct one and, as before, we estimate β_j from (1) and use (8) with R_{ft}. If we follow the steps of (12) we get

$$\bar{\tilde{v}}_{t+i}=(1-\bar{\beta})\left[(\bar{\gamma}_{0t+i}-\bar{R}_{ft+i})-\hat{\lambda}_{\gamma}(\bar{R}_{mt+i}-\bar{R}_{ft+i})\right]+\bar{e}_{t+i} \tag{13a}$$

Even if the different market factors are not correlated such that $\hat{\lambda}_f=0$ and $\hat{\lambda}_\gamma=0$ we still have a bias when we use a specific two-factor model when another two-factor-model is the correct specification. The bias is of the form

$$(1-\bar{\beta})(\bar{\gamma}_{0t+i}-\bar{R}_{ft+i}) \tag{13b}$$

Since, for example, $\bar{\gamma}_{0t}>R_{ft}$ the bias is positive for $\bar{\beta}<1$ (e.g. stock splits). This bias is not present if the single factor model is used as the incorrect model because $\hat{\alpha}$ contains the bias as can be seen from (9b). Moreover, as already mentioned, even when $\hat{\lambda}_\gamma \neq 0$ using $\hat{\alpha}$ and $\hat{\beta}$ eliminates the bias in the residuals due to the cancelling bias in $\hat{\alpha}$ and $\hat{\beta}$.

D. *Including an Irrelevant Variable*

Generally the inclusion of an irrelevant variable presents trivial problems (like inefficiency) in estimation and prediction. It mainly presents problems in small samples where efficiency of the estimate is an important property. If we assume that (1) is the correct model[13] but we use (2) with no constraints on the parameters and $\tilde{R}_k$ is the irrelevant variable then $\hat{b}_j$, from (2), is a consistent estimate of β_j. In our context where the more specific version of (2) (i.e., (8)) is used the simple problem of including an irrelevant variable turns into a more complicated problem of imposing erroneous constraints.[14]

Assuming erroneously that (8) is the true model we use the strictly constrained regression with the pair of variables; $\tilde{R}_{jt}-\tilde{R}_{kt}$ and $\tilde{R}_{mt}-\tilde{R}_{kt}$. Since we restrict the regression through the origin,[15] the estimate β_c (c denotes "constrained") is given by

$$\hat{\beta}_{jc}=\frac{\hat{E}\left[(\tilde{R}_{jt}-\tilde{R}_{kt})\cdot(\tilde{R}_{mt}-\tilde{R}_{kt})\right]}{\hat{E}(\tilde{R}_{mt}-\tilde{R}_{kt})^2} \tag{14a}$$

The bias in $\hat{\beta}_{jc}$ is therefore

$$\hat{\beta}_{jc}=\beta_j+\frac{\hat{\beta}_{jc}\hat{\sigma}^2(\tilde{R}_k)-\bar{R}_m\bar{R}_k+(\bar{R}_m-\bar{R}_k)(\alpha_j+\hat{\beta}_{jc}\bar{R}_k)}{\hat{\sigma}^2(\tilde{R}_m)+\bar{R}_m(\bar{R}_m-\bar{R}_k)} \tag{14c}$$

13. Assuming (1) to be different from (8) means that $\alpha_j \neq (1-\beta_j)E(\tilde{R}_k)$.

14. See Theil [10], pp. 545–546.

15. In empirical tests (Brenner [2]), we also tried a non-constrained version. The differences were minute.

The two bias terms on the RHS of (14c) can be trivially small or can be large depending on the relative magnitudes in (14c). As far as tests of the EMH are concerned the aggregation across time and securities may eliminate the effect of the baised $\hat{\beta}_{jc}$.

III. THE EFFECT OF CHANGES IN β; GENERAL[16]

To estimate β_j we fit a market model to the data, by OLS. This procedure assumes that β_j is constant over all t in the estimation period.[17] Using β_j, obtained from the estimation period, to get the residuals in the prediction period (denoted by i) assumes that β_j is the same constant in the prediction period too. What is the effect of such a change on tests of the EMH?

To simplify the analysis we shall assume that β_j is constant over t and over i but changes from t to i (i.e. β_j changes from the estimation period to the prediction period). Denoting[18] β_j in t as β_{jt} and β_j in i as β_{jt+i} we have

$$\beta_{jt} = \beta_{jt+i} + \zeta_j \tag{15}$$

Assuming we use the correct specification of the market model but with β_{jt} we obtain the residual

$$\begin{aligned}\hat{e}_{jt+i} &= \beta_{jt+i} R_{mt+i} + (1-\beta_{jt+i}) R_{kt+i} + e_{jt+i} - \beta_{jt} R_{mt+i} - (1-\beta_{jt}) R_{kt+i} \\ &= R_{mt+i}(\beta_{jt+i} - \beta_{jt}) - R_{kt+i}(\beta_{jt+i} - \beta_{jt}) + e_{jt+i} \\ &= (R_{mt+i} - R_{kt+i})\zeta + e_{jt+i} \qquad (16a)\end{aligned}$$

Averaging across j we get[19]

$$\bar{\hat{e}}_{t+i} = (\bar{R}_{mt+i} - \bar{R}_{kt+i})\bar{\zeta} + \bar{e}_{t+i} \tag{16b}$$

The average residual—$\bar{e}_{t+i}$—will be unbiased if the changes in β_{jt} are independent across stocks.[20] If however the change in β_{jt} is associated with the event then the changes across stocks should be in the same direction (if the event studied has a similar effect on all stocks).

Thus, an observed pattern of the residual may simply be due to a risk change and is no evidence for or against efficient markets.

16. This brief section does not deal with the problem of changes in β per se. In this section we only point out, under very general conditions, the potential effects of ignoring β changes on tests of the EMH.

17. A detailed analysis on the nonconstancy of the market parameters over time appears in Gonedes [4]. His analysis is limited to the case that parameter changes are unknown and infrequent. If the parameters change every month, any T would be as good. There would be no advantage to increasing T.

18. For simplicity we present the analysis in terms of the true β_j and not the estimate $\hat{\beta}_j$. Under the stated assumptions, however, the analysis with $\hat{\beta}_j$ is exactly the same.

19. Again, we assume $\hat{\text{cov}}(\delta_j, R_{mt+i} - R_{kt+i}) = 0$.

20. $\bar{\hat{e}}_{t+i} = \bar{e}_{t+i}$ also when $\bar{R}_{mt+i} = \bar{R}_{kt+i}$ but in general if $\bar{R}_k$ represents R_f or $\tilde{\gamma}_0$ then $\bar{R}_m > \bar{R}_k$ and the bias in $\bar{\hat{e}}_{t+i}$ will not vanish.

IV. Summary

This paper is concerned with the effect of model misspecification on tests of the efficient market hypothesis. Such misspecifications may bias the test statistics (the average residual). The bias in the residuals will, in general, originate from two sources: 1. biased parameter estimates (e.g., $\hat{\beta}$) obtained in the first stage and; 2. improper market factors used in the second stage (e.g., using $\hat{\gamma}_{0t}$ rather than R_{ft}). The effect of the biases on the average residual will depend largely on the cross-sectional distribution of β_j in the study at hand and on the time-series correlation amongst the various market factors.

The analysis here has implications for past and future research on market efficiency. Since results of past studies may have been affected by serious misspecifications, these should be considered as conditional evidence only. For future research we recommend using several alternative models to address the issue of robustness in test of the EMH.

REFERENCES

1. R. Ball. "Changes in Accounting Techniques and Stock Prices," unpublished Ph.D. Dissertation, University of Chicago, 1972.
2. M. Brenner. "The Sensitivity of the Efficient Market Hypothesis to Alternative Specification of the Market Model", unpublished Ph.D. Dissertation, Cornell University, 1974.
3. E. G. Fama, L. Fisher, M. C. Jensen and R. Roll. "The Adjustment of Stock Prices to New Information," *International Economic Review*, February 1969.
4. N. J. Gonedes. "Evidence on the Information Content of Accounting Numbers; Accounting Based and Market Based Estimates of Systematic Risk," *Journal of Financial and Quantitative Analysis*, June, 1973.
5. J. F. Jaffe. "Special Information and Inside Trading", *Journal of Business*, July 1974.
6. R. S. Kaplan and R. Roll. "Investor Evaluation of Accounting Information: Some Empirical Evidence", *Journal of Business*, April 1972.
7. G. Mandelker. "Risk and Return: The Case of Merging Firms," *Journal of Financial Economics*, December 1974.
8. M. H. Miller and M. S. Scholes. "Rates of Return in Relation to Risk: A Re-examination of Some Recent Findings", in M. C. Jensen, ed., *Studies in The Theory of Capital Markets*, New York, Praeger Publishers, 1973.
9. R. Roll. "Bias in Fitting the Sharpe Model to Time Series Data," *Journal of Financial and Quantitative Analysis*, September 1969.
10. H. Theil. *Principles of Econometrics*, New York: John Wiley & Sons, Inc. 1971.
11. E. G. Fama and J. D. Macbeth. "Risk, Return, and Equilibrium: Empirical Tests", *Journal of Political Economy*, May/June 1973.