2009

Three Bayesian econometric studies on forecast evaluation

Jingtao Wu
Iowa State University

Three Bayesian econometric studies on forecast evaluation

by

Jingtao Wu

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Economics

Program of Study Committee:
Sergio H. Lence, Major Professor
Barry L. Falk
Dermot J. Hayes
Joseph A. Herriges
Travis R.A. Sapp
Justin L. Tobias

Iowa State University

Ames, Iowa

2009

TABLE OF CONTENTS

ABSTRACT

A complete theory for evaluating forecasts has not been worked out to this date. Many studies on forecast evaluation implicitly relied on assumptions that are not supported by data, e.g., the assumption of homoskedastic and uncorrelated errors, forecaster homogeneity, etc. In this dissertation, I apply Bayesian methods to analyze various aspects of forecast evaluation. The overall objective is to better evaluate forecasts in terms of bias, efficiency, and information content by accounting for the structure of forecasts and directly addressing various critical econometric issues that are ignored by previous studies. Three related studies have been undertaken to address three issues. My first paper studies forecasts' bias and inefficiency after accounting for forecast error correlations. My second paper studies forecasts' bias and inefficiency after accounting for forecasts' hierarchical structure. My third paper proposes new measures of forecasts' information content of actual variables. Although the three papers in this dissertation studies specific data sets, the employed methods could be easily applied to forecasts with similar structures.

CHAPTER 1. GENERAL INTRODUCTION

Producing and evaluating forecasts are two major areas in economic forecasting research. Since there are usually multiple forecasts for a single variable and forecasters generally do not make their forecasting processes available, properly evaluating available forecasts is as important as, if not more than, making good forecasts. A complete theory for evaluating forecasts has not been worked out to this date. Many studies on forecast evaluation implicitly relied on assumptions that are not supported by data, such as homoskedastic and uncorrelated errors or forecaster homogeneity. Consequently, this dissertation attempts to improve the forecast evaluation literature by applying Bayesian methods that rely on more realistic assumptions. Moreover, the employed Bayesian methods, in certain cases, could yield results that are not attainable via Frequentist approaches typically used in previous studies. Specifically, three studies have been undertaken to address three different aspects of forecast evaluation. The first study, entitled "A Three-way Random Effects Study of Analyst Bias and Inefficiency," addresses the correlation structure of stock analysts' earnings forecast errors. Previous studies either did not account for forecast error correlations, or imposed assumptions not supported by the data. I deal with forecast error correlations by adding analyst, firm, and time random effects, which allow forecast errors to be correlated not only across firms and analysts, but also across time periods. I show that the three-way random effects model is a more appropriate model as the variances of all three random effects are significantly larger than zero, which contradicts the usual assumption that forecast errors are not correlated. This study also offers additional evidence on the irrationality of analyst forecasts and reconciles contradicting results in the previous literature. It also corrects an inconsistency in one previous study and classifies variables by analysts' available information set. The proposed Bayesian estimation of the three-way random effects model also offers an analysis of individual random effects, which is very difficult, if not impossible, to implement using non-Bayesian methods.

The second study, entitled “A Bayesian Hierarchical Study of Analyst Bias and Inefficiency,” relaxes the unrealistic assumption of forecaster homogeneity when studying pooled forecasts. It uses a Bayesian hierarchical model. It does not treat analysts as a homogenous population as the previous literature implicitly did by pooling forecasts across time periods and firms. The reasons to treat analyst as heterogeneous are twofold. From a theoretical point, existing studies have identified factors that affect forecast accuracy, which suggests that analysts are likely to have different forecast accuracies. Therefore, it is reasonable to allow for analysts to have different abilities, leading to different degrees of bias and inefficiency in their forecasts. From a practical standpoint, investors appear to think that analysts have different skills. The main contribution of the present study is that, as a result of the proposed method, I am able to assess forecasts at the individual analyst level. This allows me to identify that there is heterogeneity in the degrees of analysts' bias and inefficiency. The forecasts of some analysts, especially those of some firms that these analysts follow, are found to be unbiased and efficient.

The third study, entitled “The Information Content of Forecasts,” proposes a new approach to compare forecasts’ information contents. Previous studies’ regression-based measures are prone to multicollinearity problems as forecasts for the same variable are usually highly correlated. By regarding forecasts as predictors, I derive next period’s expected value of the variable being forecasted conditional on alternative information sets using the Kalman filter. Forecasts that contain more information will lead to a smaller variance of deviations between actual values and expected values. The relative magnitude of the above variance is regarded as the measure of the relative information contents of competing forecasts. I also propose a way to measure the information content of forecasts from the same source without competing forecasts, which could not be determined by previous methods. The advocated measures are computed for a well-known data set and yield different, yet compelling, conclusions from those drawn by the previous literature's regression-based measures.

CHAPTER 2. A THREE-WAY RANDOM EFFECTS STUDY OF ANALYST BIAS AND INEFFICIENCY

2.1 Abstract

Using a three-way random effects model, consistent with the previous literature, the present study shows that as a population, analysts' forecasts are biased and inefficient. The present study offers additional evidence regarding the irrationality of analyst forecasts and reconciles previous contradicting results. In contrast to Keane and Runkle (1998), it is shown that the null hypothesis of analyst forecast rationality is rejected even if forecast error correlations and special charges are accounted for. The present study also corrects an inconsistency in Constantinou, Forbes and Skerratt (2003) and classifies variables by analysts' available information set. This re-classification is shown to significantly change the results previously reported by the literature. The present study contributes to the literature on stock analysts by correctly accounting for forecast error correlations. Previous studies either did not account for forecast error correlations or adopted questionable assumptions (e.g., Keane and Runkle (1998)). In addition, evidence is provided about analysts' asymmetric responses to different types of past information.

2.2 Introduction

Stock analysts who make forecasts of earnings per share (EPS) are an interesting group of economic agents to study for several reasons. First, researchers have found that earnings forecasts appear to have economic value for investors (e.g., Womack (1996)). Second, analysts' forecasts have often been found to outperform time series models, suggesting that analysts are rather good at what they do (e.g., Brown, Hagerman, Griffin, and Zmijewski (1987)). Third, the precision of analysts' forecasts represents an upper bound of the quality of earnings forecasts made by less sophisticated agents.

Although the importance of analysts is beyond dispute, the quality of their forecasts has become the subject of intense research and debate. An important category of research on analysts' forecasts examines their bias and inefficiency. In this literature, forecasts are usually considered to be biased if there is a systematic positive (or negative) difference between the forecasts and the actual EPS, whereas forecasts are typically labeled inefficient if they do not fully incorporate past information available at the time of issuing the forecasts. One of the most widely held beliefs in the literature is that analysts produce biased forecasts that are "too optimistic" (see, e.g., review by Kothari (2001)). In addition, numerous studies have documented analysts' inefficiency with respect to public information such as past earnings levels (e.g., De Bondt and Thaler (1990)), past earnings changes (e.g., Abarbanell and Bernard (1992)), extreme past earnings changes (e.g., Easterwood and Nutt (1999)), past returns (e.g., Lys and Sohn (1990), Abarbanell (1991), Ali, Klein and Rosenfeld (1992)), past forecast errors (e.g., Ali, Klein and Rosenfeld (1992)), and past forecast revisions (e.g., Amir and Ganzach (1998)), Mendenhall (1991)). Studies finding evidence of systematic under-reaction include Lys and Sohn (1990), Abarbanell (1991), Abarbanell and Bernard (1992), Ali, Klein and Rosenfeld (1992), and Elliot, Philbrick and Wiedman (1995). In contrast, De Bondt and Thaler (1990) find that changes in EPS forecasts are too extreme, consistent with systematic over-reaction. Easterwood and Nutt (1999) attempted to reconcile the conflicting evidence by testing the robustness of Abarbanell and Bernard (1992) under-reaction results to the nature of the information. They find that financial analysts under-react to negative information but over-react to positive information. Ali, Klein and Rosenfeld (1992), among others, document that analysts' forecasts are inefficient with respect to their most recent forecast error.

Another category of research focuses on the properties of individual analysts' forecasts, by studying the determinants of their accuracy (Mikhail, Walther and Willis (1997), Jacob, Lys and Neale (1999), Clement (1999)). These studies suggest that experience, the size of the brokerage firm that an analyst works for, and the number of firms and industries followed by an analyst affect forecast accuracy.

Since each firm is usually followed by several analysts, there are typically multiple EPS forecasts given by different analysts for a given firm and period. On the other hand, each analyst usually follows several firms at the same time. Therefore, forecasts can be grouped by firms, analysts, and time periods. The existence of the above data hierarchies is neither accidental nor ignorable. Once a group (a specific firm, a specific analyst, or a specific time period) is established, it often tends to become differentiated from the other groups. This differentiation implies that the group and its members both influence and are influenced by the group membership. To ignore this relationship risks overlooking the importance of group effects, and it may also render invalid many of the traditional statistical techniques used for studying data relationships (Goldstein (2003)). Several authors (e.g., Crichfield, Dyckman and Lakonishok (1978), O'Brien (1988), Abarbanell (1991), Abarbanell and Bernard (1992)) have noted that statistical inference about the properties of analysts' forecasts is very difficult if forecast errors are correlated across analysts or firms.

Some studies have attempted to deal with the problem of correlated errors across analysts/firms. Crichfield, Dyckman and Lakonishok (1978) first noted that forecasts for all companies may be cross-sectionally correlated due to aggregate market events and suggested that a relatively long time span is required to test the ability of analysts. O'Brien (1988) first allowed for random time-specific shocks to deal with aggregate shocks and found weak evidence that forecasts are upward-biased. Keane and Runkle (1998) extended O'Brien (1988) by allowing for firm-specific as well as analyst-specific shocks. However, Keane and Runkle (1998) assumed that forecast errors are not correlated across time.

The present study further extends Keane and Runkle (1998)'s work by allowing forecast errors to be correlated not only across firms and analysts, but also across time periods. The three-way random effects model used here is a form of hierarchical model where random effects are assumed to be normally distributed among analysts, firms, and time periods. Therefore, the random effects have the form of a shrinkage estimator, utilizing the complete sample information in addition to the group's information when estimating the random effects of each group. The three-way random effects model is estimated by Gibbs sampling. Using Gibbs sampling enables any posterior joint or marginal distribution of interest to be constructed, in principle to any degree of accuracy. Moreover, having available full posterior distributions instead of normal approximations can be valuable, particularly for highly skewed posteriors where maximum likelihood estimates are misleading. The Gibbs sampling approach also allows posterior distributions to be easily calculated for arbitrary functions of parameters.

This present study shows that the three-way random effects model is a more appropriate model for the analyst forecast quality problem than Keane and Runkle (1998)’s model. This is true because the variances of all three random effects are significantly larger than zero, which contradicts Keane and Runkle (1998)’s assumption that forecast errors are not auto-correlated. The present study contributes to the research on stock analysts by correctly accounting for forecast error correlations. It offers additional evidence on the irrationality of analyst forecasts and reconciles previous contradicting results. In contrast to Keane and Runkle (1998), it is shown that the null hypothesis of analyst forecast rationality is rejected even if forecast error correlations and special charges are accounted for. The present study also corrects an inconsistency made by Constantinou, Forbes and Skerratt (2003) and classifies variables by analysts’ available information set. The proposed Bayesian estimation of the three-way random effects model also offers an analysis of individual random effects, which is very difficult, if not impossible, to implement using non-Bayesian methods.

The rest of the paper is organized as follows. Section 3 introduces the three-way random effects model and outlines the Gibbs sampling method. Section 4 describes the data. Section 5 discusses results regarding analysts' bias and inefficiency. Section 6 concludes.

2.3 Estimation Framework

This section outlines the three-way random effects model and the Gibbs sampling method used to estimate it. The tests for forecasts' bias and inefficiency commonly used in the literature are also described.

2.3.1 Testing for bias

Let r_{t+1} denote the actual value of variable r at time $t+1$ and let x_t^{t+1} denote the forecast of r_{t+1} as of time t. Many, perhaps most, empirical studies of survey forecasts test for bias using the Mincer-Zarnowitz regression (Mincer and Zarnowitz (1969)) of the form,

$$r_{t+1} = \alpha + \beta x_t^{t+1} + \varepsilon_{t+1}, \tag{2.1}$$

where α and β are regression coefficients and ε is a regression residual. Given (2.1), rejecting the null hypothesis $H0:(\alpha,\beta)=(0,1)$ provides evidence of bias in the forecasts.[1]

If one imposes $\beta = 1$ and subtracts x_t^{t+1} from both sides of equation (2.1), the transformed regression is,

[1] If r_{t+1} and x_t^{t+1} are nonstationary, both series are usually differenced by lagged actual values, transforming the levels regression into the "returns" regression $(r_{t+1} - r_t) = \alpha + \beta(x_t^{t+1} - r_t) + \varepsilon_{t+1}$. In this specification of nonstationary variables, unbiasedness is still tested using the same null hypothesis $H0:(\alpha,\beta)=(0,1)$ as in the levels regression.

$$(r_{t+1} - x_t^{t+1}) = \alpha + \varepsilon_{t+1} . \tag{2.2}$$

Testing that $\alpha = 0$ in the last regression is equivalent to jointly testing that $\alpha = 0$ and $\beta = 1$ in the Mincer-Zarnowitz regression. In the transformed regression, the null hypothesis $H0 : \alpha = 0$ is equivalent to having forecast errors with no systematic bias. The idea is that unbiased forecasts should not be systematically smaller or larger than actual values. Some researchers (e.g., Carmona (2005)) prefer equation (2.2) over equation (2.1) for several reasons. First, equation (2.2) is more parsimonious. Second, in equation (2.1), x_t^{t+1} is required to be uncorrelated with ε_{t+1} for estimates to be consistent. The above condition may not necessarily hold. Finally, if r_{t+1} and x_t^{t+1} are persistent (or unit-root) processes, then the normal distribution may be a poor approximation to the distribution of the standard test of $H0 : (\alpha, \beta) = (0,1)$. Because of the advantages of equation (2.2) over (2.1), equation (2.2) will be used to analyze forecasts in the present study.

2.3.2 Inclusion of multiple random effects

Suppose at time t, analyst a follows firm f . The realization of firm f 's next quarter EPS, E_{ft+1}, is unknown to analyst a .[2] The forecast of E_{ft+1} by a as of time t is designated by F_{aft}^{t+1} . Defining the forecast error of analyst a for firm f 's EPS from t to $t+1$ as $y_{aft+1} = E_{ft+1} - F_{aft}^{t+1}$, equation (2.2) can be written as,

$$y_{aft+1} = \alpha + error_{aft+1} . \tag{2.3}$$

The term $error_{aft+1}$ may have a complex covariance structure due to the following three types of correlations. For a given analyst a , $error_{aft+1}$ may be correlated across firms and time because of his personal ability, foresight, forecasting methods used, etc. For example, he may be pessimistic and have a tendency to under-estimate EPS for all firms he follows. For a given firm f , $error_{aft+1}$ may be correlated across analysts and time because of firm-specific

---

[2] There is not a subscript a for E_{ft+1}, because firm f 's actual EPS does not depend on analyst a .

characteristics. For example, one firm could be rapidly growing and consistently beating analysts' forecasts. Finally, for a given time period t, $error_{aft+1}$ may be correlated across analysts and firms because of unanticipated systematic shocks to the economy.

The present study allows for forecast error correlations across analysts, firms, and time periods by extending equation (2.3) to include analyst-specific, firm-specific and time-specific random effects as in equation (2.4).

$$y_{aft+1} = \alpha + \mu_a + \mu_f + \mu_{t+1} + \varepsilon_{aft+1}. \tag{2.4}$$

In equation (2.4), α is the average of all forecast errors, μ_a is the deviation of the average of analyst a's forecast errors from the average of the sample, μ_f is the deviation of firm f's forecast errors from the average of the sample, and μ_{t+1} is the deviation of the average of period $t+1$'s forecast errors from the average of the sample. The idiosyncratic error ε_{aft+1} is assumed to be iid $N(0,\sigma^2)$. To justify the assumption that the idiosyncratic error is homoskedastic, in the present empirical application, the actual EPS and forecasts are scaled by EPS standard deviations. This normalization procedure is crucial to justify the assumption of homoskedasticity and is common practice in previous studies. The analyst random effects are μ_a and are assumed to be iid $N(0,\sigma_a^2)$ for $a = 1,\cdots,A$. Similarly, the firm random effects are μ_f and are assumed to be iid $N(0,\sigma_f^2)$ for $f = 1,\cdots,F$ and the time random effects are μ_t and are assumed to be iid $N(0,\sigma_t^2)$ for $t = 1,\cdots,T$.

Equation (2.4) is a three-way random effects model that can account for the complex patterns of correlation across forecast errors. To see the resulting covariance structure of $error_{aft+1}$ induced by equation (2.4), suppose that there are two analysts ($a1, a2$) who give forecasts for two firms ($f1, f2$) in two periods ($t1, t2$). In the matrix form, equation (2.4) for this case will be,

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \alpha + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{a1} \\ \mu_{a2} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{f1} \\ \mu_{f2} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{t1} \\ \mu_{t2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{121} \\ \varepsilon_{122} \\ \varepsilon_{211} \\ \varepsilon_{212} \\ \varepsilon_{221} \\ \varepsilon_{222} \end{bmatrix}.$$

Therefore, the $error_{aft+1}$ in equation (2.3) can be decomposed as,

$$error \equiv \begin{bmatrix} error_{111} \\ error_{112} \\ error_{121} \\ error_{122} \\ error_{211} \\ error_{212} \\ error_{221} \\ error_{222} \end{bmatrix} = \begin{bmatrix} \mu_{a1} + \mu_{f1} + \mu_{t1} + \varepsilon_{111} \\ \mu_{a1} + \mu_{f1} + \mu_{t2} + \varepsilon_{112} \\ \mu_{a1} + \mu_{f2} + \mu_{t1} + \varepsilon_{121} \\ \mu_{a1} + \mu_{f2} + \mu_{t2} + \varepsilon_{122} \\ \mu_{a2} + \mu_{f1} + \mu_{t1} + \varepsilon_{211} \\ \mu_{a2} + \mu_{f1} + \mu_{t2} + \varepsilon_{212} \\ \mu_{a2} + \mu_{f2} + \mu_{t1} + \varepsilon_{221} \\ \mu_{a2} + \mu_{f2} + \mu_{t2} + \varepsilon_{222} \end{bmatrix}.$$

Let $\sigma_{aft}^2 \equiv \sigma_a^2 + \sigma_f^2 + \sigma_t^2$, $\sigma_{at}^2 \equiv \sigma_a^2 + \sigma_t^2$, $\sigma_{ft}^2 \equiv \sigma_f^2 + \sigma_t^2$, the covariance matrix of $error$, $\text{cov}(error, error^{'})$, is:

$$\begin{bmatrix}
\sigma_{aft}^2+\sigma^2 & \sigma_{af}^2 & \sigma_{at}^2 & \sigma_a^2 & \sigma_{ft}^2 & \sigma_f^2 & \sigma_t^2 & 0 \\
\sigma_{af}^2 & \sigma_{aft}^2+\sigma^2 & \sigma_a^2 & \sigma_{at}^2 & \sigma_f^2 & \sigma_{ft}^2 & 0 & \sigma_t^2 \\
\sigma_{at}^2 & \sigma_a^2 & \sigma_{aft}^2+\sigma^2 & \sigma_{af}^2 & \sigma_t^2 & 0 & \sigma_{ft}^2 & \sigma_f^2 \\
\sigma_a^2 & \sigma_{at}^2 & \sigma_{af}^2 & \sigma_{aft}^2+\sigma^2 & 0 & \sigma_t^2 & \sigma_f^2 & \sigma_{ft}^2 \\
\sigma_{ft}^2 & \sigma_f^2 & \sigma_t^2 & 0 & \sigma_{aft}^2+\sigma^2 & \sigma_{af}^2 & \sigma_{at}^2 & \sigma_a^2 \\
\sigma_f^2 & \sigma_{ft}^2 & 0 & \sigma_t^2 & \sigma_{af}^2 & \sigma_{aft}^2+\sigma^2 & \sigma_a^2 & \sigma_{at}^2 \\
\sigma_t^2 & 0 & \sigma_{ft}^2 & \sigma_f^2 & \sigma_{at}^2 & \sigma_a^2 & \sigma_{aft}^2+\sigma^2 & \sigma_{af}^2 \\
0 & \sigma_t^2 & \sigma_f^2 & \sigma_{ft}^2 & \sigma_a^2 & \sigma_{at}^2 & \sigma_{af}^2 & \sigma_{aft}^2+\sigma^2
\end{bmatrix}.$$

The most comprehensive study attempting to deal with the problem of correlated forecast errors is Keane and Runkle (1998). The forecast error covariances that Keane and Runkle (1998) assumed for $error$ are the following,

$$\begin{bmatrix} a & 0 & b & 0 & c & 0 & d & 0 \\ 0 & a & 0 & b & 0 & c & 0 & d \\ b & 0 & a & 0 & b & 0 & c & 0 \\ 0 & b & 0 & a & 0 & b & 0 & c \\ c & 0 & d & 0 & a & 0 & b & 0 \\ 0 & c & 0 & d & 0 & a & 0 & b \\ d & 0 & c & 0 & b & 0 & a & 0 \\ 0 & d & 0 & c & 0 & b & 0 & a \end{bmatrix} \tag{2.5}$$

where a, c, b, d are constants to be estimated. The main difference between the forecast error covariance proposed here and that of Keane and Runkle (1998) is that forecast errors of different time periods are not correlated. In the present study, only the forecast errors of different analysts for different firms are not correlated across time periods, due to the lack of common random effects.

The covariance structure of the present study represents an improvement because of the following two reasons. First, whether forecast errors are correlated across time periods or not depends on the estimated random effects variances. Comparing the two sets of covariance matrices, we have $a = \sigma_a^2 + \sigma_f^2 + \sigma_t^2 + \sigma^2$, $b = \sigma_a^2 + \sigma_t^2$, $c = \sigma_f^2 + \sigma_t^2$, $d = \sigma_t^2$. Since (2.5) restricted a, c, b, d to be of different values, σ_a^2 and σ_f^2 should be non-zeros values, which indicates that the present study's covariance structure is more suitable even under Keane and Runkle's reasoning. Second, previous research has documented that forecast errors are positively auto-correlated (e.g., Ali, Klein and Rosenfeld (1992)). As shown later, the present study also documented positive auto-correlation in analysts' forecast error. Therefore, the covariance matrix (2.5) is not supported by data.

2.3.3 Testing for efficiency

Because an efficient forecast incorporates all available information, public and private, it follows that there should be no relationship between forecast errors and any variable known to analysts at the time of the forecast. To test for efficiency, all variables known at t should be included on the right-hand side of equation (2.4),

$$y_{aft+1} = \alpha + \beta X_{aft} + \mu_a + \mu_f + \mu_{t+1} + \varepsilon_{aft+1},$$

where X_{aft} are variables known at t. Since it is not feasible to include all variables known to analysts at the time of their forecasts, only a weak efficiency condition, i.e., whether forecasts are efficient with respect to a small set of relevant variables, is tested following the previous literature. Based on previous studies, two obvious variables to test the weak efficiency condition are past quarter forecast errors $(PQFE = E_{ft} - F^{t}_{aft-1})$ and past quarter EPS changes $(PQEC = E_{ft} - E_{ft-1})$. Therefore, equation (2.4) is extended as,

$$y_{aft+1} = \alpha + \beta_1 PQFE + \beta_2 PQEC + \mu_a + \mu_f + \mu_{t+1} + \varepsilon_{aft+1}. \tag{2.6}$$

If E^{t+1}_{aft} is an efficient forecast, both β_1 and β_2 will be zero. Non-zero values of β_1 and β_2 means that analysts could improve their forecasts by incorporating information that is available at the time of the forecast.

2.3.4 Modeling analysts' under-reaction and over-reaction

In addition to inefficiency, Abarbanell and Bernard (1992) gave a second interpretation of the sign of β_2. They suggested that $\beta_2 > 0\ (< 0)$ indicates under-reaction (over-reaction) to the prior earnings changes. Their interpretation is the following. For the case of $\beta_2 > 0$ and $PQEC > 0\ (< 0)$, equation (2.5) specifies $\beta_2 PQEC > 0\ (< 0)$ which implies $y_{aft+1} = E_{ft+1} - F^{t+1}_{aft} > 0\ (< 0)$, excluding other terms in the equation. The under-reaction explanation is that analysts are too cautious about current EPS rising (decreasing) further and give a forecast that is too low (high). Therefore, too little weight is given by the analyst to $PQEC$.

Constantinou, Forbes, and Skerratt (2003) (henceforth CFS) pointed out that the interpretation of β_2 depends on whether earnings follow a trend (momentum) or earnings are mean-reverting (reversion). CFS classified earnings momentum and reversions by the signs of $PQEC$ and current quarter earnings change $CQEC$ (defined as $E_{t+1} - E_t$). If $PQEC$ and $CQEC$ have the same signs, CFS classified $PQEC$ as a momentum case. If $PQEC$ and $CQEC$ are opposite in signs, CFS classified $PQEC$ as a reversion case. The interpretation given by Abarbanell and Bernard (1992) essentially refers to the earnings momentum cases. In the cases of earnings reversions, when $\beta_2 > 0$ and $PQEC > 0$, it implies that analysts under-estimate EPS. Since $PQEC > 0$, the reversion in earnings implies current quarter EPS will decrease. Therefore, analysts over-react to the information contained in $PQEC$ and give a forecast that is too low warranted by mean reversion. CFS found substantial under-reaction, particularly in situations of earnings momentum.

However, the classification in CFS is problematic in the sense that analysts obviously do not know the sign of $CQEC$ when they make forecasts of E_{t+1}. Therefore, analysts would not be able to know whether earnings are in the momentum regime or reversion regime by the CFS definition. To include explanatory variables that are in the analysts' information set at the time of issuing the forecast, the present study classifies earnings momentum and reversions by the signs of $PQEC$ and its first lag, $lagPQEC$ (defined as $E_{ft-1} - E_{ft-2}$). If $PQEC$ and $lagPQEC$ have the same signs, $PQEC$ is classified as a momentum case, $PQEC_M$. If $PQEC$ and $lagPQEC$ are opposite in signs, $PQEC$ is classified as a reversion case, $PQEC_R$. Equation (2.6) can be extended as,

$$y_{aft+1} = \alpha + \beta_1 PQFE + \beta_2^R PQEC_R + \beta_2^M PQEC_M + \mu_a + \mu_f + \mu_{t+1} + \varepsilon_{aft+1} \tag{2.7}$$

Further extensions of equation (2.7) can be used to distinguish different cases of $PQFE$, $PQEC_R$, and $PQEC_M$ so as to test for analysts' asymmetric responses. $PQFE$ can be divided into positive errors ($PQFE_P$) and negative errors ($PQFE_N$). $PQFE_P$ equals $PQFE$ when $PQFE$ is positive (analysts under-estimate last quarter's EPS) and equals zero otherwise. $PQFE_N$ equals $PQFE$ when $PQFE$ is negative (analysts over-estimate last quarter's EPS) and equals zero otherwise. $PQEC_R$ can be divided into upward reversion ($PQEC_{UR}$) and downward reversion ($PQEC_{DR}$). $PQEC_{UR}$ equals $PQEC_R$ when

$PQEC_R$ is positive and equals zero otherwise. $PQEC_{DR}$ equals $PQEC_R$ when $PQEC_R$ is negative and equals zero otherwise. $PQEC_M$ can be divided into upward momentum ($PQEC_{UM}$) and downward reversion ($PQEC_{DM}$). $PQEC_{UM}$ equals $PQEC_M$ when $PQEC_M$ is positive and equals zero otherwise. $PQEC_{DM}$ equals $PQEC_M$ when $PQEC_M$ is negative and equals zero otherwise. Equation (2.8) extends equation (2.7) to include all these cases:

$$\begin{aligned} y_{aft+1} = \alpha + \beta_1^P PQFE_P + \beta_1^N PQFE_N + \beta_2^{UR} PQEC_{UR} + \beta_2^{DR} PQEC_{DR} \\ + \beta_2^{UM} PQEC_{UM} + \beta_2^{DM} PQEC_{DM} + \mu_a + \mu_f + \mu_{t+1} + \varepsilon_{aft+1} \end{aligned} \tag{2.8}$$

2.3.5 Estimation method: Gibbs sampling

The parameters in equations (2.4), (2.6), (2.7), and (2.8) are estimated by Gibbs sampling. Using the Gibbs sampler, the joint posteriors of all parameters can be analyzed one set at a time. By cycling repeatedly through draws of each parameter conditional on the remaining parameters, the Gibbs sampler produces a Markov chain of parameter draws whose joint distribution converges to the posterior. It is assumed that the random effects are independently and normally distributed with zero means and certain variances which themselves have inverse-Gamma distributions with corresponding hyper-parameters. This subsection outlines the priors employed for the parameters and the joint likelihood function. The complete Gibbs sampler with conditional posterior distributions for each set of parameters is given in the Appendix.

The joint posterior distribution for parameters of the proposed model for equation (2.4) can be written as,[3]

---

[3] Since it is straight-forward to extend the Gibbs sampling to include more independent variables to accommodate the estimation of equations (2.5) to (2.7), for simplicity, the notation in this section and the Appendix only refers to equation (2.4).

$$p(\Gamma \mid data) \propto \prod_{aft}^{N} \Phi(y_{aft+1} \mid \alpha, \mu_a, \mu_f, \mu_{t+1}, \sigma^2) \prod_{a}^{A} \Phi(\mu_a \mid \sigma_a^2) \prod_{f}^{F} \Phi(\mu_f \mid \sigma_f^2) \prod_{t}^{T} \Phi(\mu_{t+1} \mid \sigma_t^2)$$
$$* \Phi(\alpha \mid \mu_\alpha, V_\alpha)\, p(\sigma^2 \mid a, b)\, p(\sigma_a^2 \mid a_a, b_a)\, p(\sigma_f^2 \mid a_f, b_f)\, p(\sigma_t^2 \mid a_t, b_t)$$

where $\Gamma = \left(\alpha, \{\mu_a\}_{a=1}^{A}, \{\mu_f\}_{f=1}^{A}, \{\mu_{t+1}\}_{t=1}^{T}, \sigma^2, \sigma_a^2, \sigma_f^2, \sigma_t^2\right)$ denotes all parameters in the model and $\Phi(\bullet)$ denote the normal probability density function. The priors employed in the Gibbs sampling are as follows:

$$\alpha \sim N(\mu_\alpha, V_\alpha)$$
$$\varepsilon_{aft+1} \overset{iid}{\sim} N(0, \sigma^2)$$
$$\mu_a \overset{iid}{\sim} N(0, \sigma_a^2), \text{ for } a = 1, \cdots, A$$
$$\mu_f \overset{iid}{\sim} N(0, \sigma_f^2), \text{ for } f = 1, \cdots, F$$
$$\mu_{t+1} \overset{iid}{\sim} N(0, \sigma_t^2), \text{ for } t = 1, \cdots, T$$
$$\sigma^2 \sim IG(a, b)$$
$$\sigma_a^2 \sim IG(a_a, b_a)$$
$$\sigma_f^2 \sim IG(a_f, b_f)$$
$$\sigma_t^2 \sim IG(a_t, b_t)$$

where $IG(\bullet)$ denotes the inverse Gamma function. For the normal prior of α, the present study uses $\mu_a = 0$ and $V_a = 1000$. The prior is chosen so that the prior mean is 0 and the prior variance is large, which makes the prior non-informative.[4] The values of inverse Gamma hyper-priors are set as follows, $a = a_a = a_f = a_t = 3$ and $b = b_a = b_f = b_t = 3$. The hyper-priors are chosen so that the prior means for the variances is 1/6. The prior means for

[4] For regressions with multiple regressors, the coefficients α and β are given multivariate normal priors with zero means and large covariance matrix. For example, in the estimation of equation (2.5), the non-informative prior given to α, β_1 and β_2 consists of $[\alpha, \beta_1, \beta_2]' \sim N(0_3, I_3 * 1000)$, where $0_3 = [0 \quad 0 \quad 0]'$ and I_3 is a 3×3 identity matrix.

variances are chosen to be of the same magnitude as the standard deviation of the scaled forecast errors. The degrees of freedom for the inverse-Gamma hyper-priors are chosen to be small so that prior distributions are spread around their means. These a and b values make the inverse Gamma priors non-informative. Estimation results remain basically unchanged if different prior values are used.

2.4 Data

The present study focuses on analysts covering the U.S. equity market from the third quarter of 1997 until the first quarter of 2005, altogether 31 quarters. The EPS measures used are quarterly. Because of time random effects, firms in the sample are restricted to have fiscal years ending in December. Following Keane and Runkle (1998), which required a minimum number of forecasts for the companies in the sample, firms in the sample are further restricted to have at least 100 quarterly forecasts. The final sample has 604 firms followed by 1122 analysts. The total number of analyst-firm-quarter observations is 121886.

Because the levels of EPS vary across firms, most studies on EPS forecasts use some sort of scaling procedure to mitigate the heteroskedasticity problem. The present study scales the EPS changes and forecast errors by the corresponding firm's standard deviations of quarterly EPS from 1994 to 1996. De Bondt and Thaler (1990) used EPS standard deviations to scale EPS, and showed that it yields qualitatively similar results as scaling with stock prices or company assets. The scaled forecast errors have a mean of -0.125 and a standard deviation of 1.309. Its histogram shows that forecast errors are slightly skewed to the left. The 5th percentile of the scaled forecast errors is -2.443 and the 95th percentile is 1.544. As for the signs of the forecast errors, 48.54% are positive (analysts under-estimate), 40.90% are negative (analysts over-estimate) and 10.55% are zero.

Since there are more positive errors than negative errors and the error mean is negative, analysts must, on average, have made bigger mistakes when they over-estimated. This seems to validate Keane and Runkle (1998)'s point that special attention should be paid to discretionary asset write-downs, as asset write-offs and other before-tax special charges negatively affect earnings. To accommodate this argument and compare with Keane and Runkle (1998)'s study, the present analysis uses four alternative approaches to eliminate the potential effects of discretionary asset write-downs.

First, if any quarter's EPS of a particular firm from third quarter of 1997 until the first quarter of 2005 is four standard deviations away form the mean EPS during that period, that firm is dropped from the sample. Denote this sample as ByEPSDev. It has 561 firms followed by 910 analysts with 79661 observations. Secondly, large forecast errors are identified. There are 99 forecast errors that are four standard deviations above the sample mean, and 1053 forecast errors that are four standard deviations below the sample mean. Three samples are derived by excluding these large forecast errors using different schemes. The first sample, ByAnalystFirm, is derived by excluding the analyst-firm combinations that contain these outliers, i.e., if one analyst made a large forecast error for a particular firm, then this analyst's forecast history for this firm is dropped from the sample. The ByAnalystFirm sample has 604 firms followed by 1114 analysts with 107802 observations. The second sample, ByFirm, is derived by excluding the firms that contain these outliers, i.e., if one firm has a large forecast error, then this firm's forecast history is dropped from the sample. The ByFirm sample has 396 firms followed by 1087 analysts with 75873 observations. The third sample, ByAnalyst, is derived by excluding the analysts that contain these outliers, i.e., if one analyst has a large forecast error, then this analyst's forecast history is dropped from the sample. The ByAnalyst sample has 587 firms followed by 580 analysts with 56338 observations.

An interesting fact is that the number of firms in the ByAnalystFirm sample is the same as in the complete sample. Excluding the analyst-firm combinations that contain outliers eliminate the entire forecast histories of 8 analysts, and part of the forecast history for some remaining analysts. This raises questions about Keane and Runkle (1998)'s point of discretionary asset write-downs. If discretionary asset write-downs are an important problem, then no analyst should be able to forecast the EPS of the firms that have discretionary asset write-downs. Therefore, these firms will be completely dropped from the sample. However, the present study did not identify any firm that is not forecastable and only identified analysts who could not forecast. Furthermore, many special charges identified by Keane and Runkle (1998) are related to corporate restructure, plant closing, etc. These events are public information and analysts should be able to forecast, at least partially, some of their impacts on EPS. Moreover, in their sensitivity analysis, Keane and Runkle used a 3.5- and a 4.5-standard-deviation cutoff, and showed the results remained the same. This indicates that the effects of special charges may not be large.

The percentages of various regressors are of interest for the efficiency analysis. They are presented below for the complete sample, with the ByAnalystFirm sample percentages shown between parentheses. The percentages of $PQEC$ classified as earnings momentum and earnings reversions are 48.45% (48.04%) and 51.55% (51.96%), respectively. The percentages of $PQEC$ that are classified as upward earnings momentum and downward earnings momentum are 33.33% (33.31%) and 15.12% (14.73%), respectively. Finally, the percentages of $PQEC$ that are classified as upward earnings reversions and downward earnings reversions are 25.75% (25.97%) and 25.80% (25.99%), respectively.

2.5 Estimation Results

2.5.1 Estimation results regarding bias

When estimating equation (2.4), the Gibbs sampler is run for 5000 iterations. The first 1000 iterations are discarded as the burn-in period. Different chains were run with different and over-dispersed starting values. The commonly used convergence tests (e.g., the Geweke diagnostic test and the Gelman and Rubin diagnostic test) are performed on the parameters $\alpha, \sigma^2, \sigma_a^2, \sigma_f^2, \sigma_t^2$ and randomly selected estimated random effects. All of the parameters tested passed the convergence tests. The trace plots show that the simulated draws from different chains seem to settle down and explore the same region very quickly. In fact, after the first ten iterations or so, the progression of different chains seems very similar. These provide suggestive evidence that the choice of the number of iterations and burn-in periods are adequate for the present application.

The estimated posterior means and standard deviations of the parameters are reported in Table 2.1. From Table 2.1, it can be concluded that the posterior standard deviations are small compared with the posterior means, indicating that the parameters are accurately measured. The posterior means of intercept α are -0.139 for the complete sample, -0.084 for the ByAnalystFirm sample, and -0.133 for the ByEPSDev sample. All of them are significantly negative. Since the dependent variable is actual EPS minus the forecast scaled by the EPS standard deviation, the negative intercept could be regarded as forecasts being too optimistic as the average error is negative. This result is consistent with the results in the previous literature as shown in the review by Kothari (2001), which finds that analysts are too optimistic.

The results from the complete sample and the ByEPSDev sample are very similar. Eliminating firms that have unusual EPS could not bring unbiasedness to EPS forecasts. This is contradictory to Keane and Runkle’s finding. However, the posterior means of α in the

ByFirm and the ByAnalyst samples are not significantly different from zero. The ByFirm sample excluded firms that have large forecast errors. The ByAnalyst sample deleted analysts that have large forecast errors. The above results indicate that dropping bad analysts or difficult-to-forecast firms could make the EPS forecasts unbiased as a population. Therefore, it is quite possible for studies involving small samples, which do not contain bad analysts or difficult-to-forecast firms, to find that forecasts are unbiased.

Table 2.1 Estimation Results Regarding EPS Forecasts' Bias

	All Post. Mean	ByAnalystFirm Post. Mean	ByFirm Post. Mean	ByAnalyst Post. Mean	ByEPSDev Post. Mean
α	-0.139 (0.045)	-0.084 (0.047)	-0.041 (0.035)	-0.037 (0.034)	-0.133 (0.059)
σ^2	1.486 (0.014)	1.007 (0.010)	0.710 (0.005)	0.872 (0.006)	0.964 (0.018)
σ_a^2	0.034 (0.002)	0.021 (0.002)	0.018 (0.001)	0.023 (0.002)	0.032 (0.002)
σ_f^2	0.149 (0.009)	0.103 (0.007)	0.069 (0.005)	0.094 (0.007)	0.132 (0.009)
σ_t^2	0.075 (0.019)	0.060 (0.015)	0.052 (0.013)	0.057 (0.014)	0.062 (0.015)
No. of Analysts	1122	1114	1087	580	910
No. of Firms	604	604	396	587	561
No. of Obs.	121886	107802	75873	56338	79661

Notes:

1. The regression specification is $y_{aft+1} = \alpha + \mu_a + \mu_f + \mu_{t+1} + \varepsilon_{aft+1}$. The dependent variable $y_{aft+1} = E_{ft+1} - F_{aft}^{t+1}$, and is the forecast error of next quarter's EPS. $\varepsilon_{aft+1} \overset{iid}{\sim} N(0, \sigma^2)$. The analyst random effects $\mu_a \overset{iid}{\sim} N(0, \sigma_a^2)$, for $a = 1, \cdots, A$. The firm random effects $\mu_f \overset{iid}{\sim} N(0, \sigma_f^2)$ for $f = 1, \cdots, F$. The time random effects $\mu_t \overset{iid}{\sim} N(0, \sigma_t^2)$, for $t = 1, \cdots, T$.

2. Posterior standard deviations are shown between parenthesis. Post. Mean refers to the mean of the parameter draws in the MCMC chain. Posterior standard deviations are the standard deviations of the parameter draws in the MCMC chain.

Figure 2.1 depicts the histograms of idiosyncratic error variance as well as the variance of random effects for the complete sample. Because all of the random effects are assumed to have zero means, the estimated variances can be seen as a measure of how significant the random effects are. The variances of the three random effects are all significantly larger than zero, which means that time periods, firms and analysts all have effects on the forecast errors. Since σ_a^2 and σ_f^2 are significantly larger than zero, the forecast errors are correlated across time as shown before. This contradicts Keane and Runkle (1998)'s assumption that forecast errors are not correlated. However, the magnitude of the idiosyncratic error variance σ^2 is 15 to 20 times larger than σ_t^2, σ_a^2 and σ_f^2, indicating that the idiosyncratic error is by far the most important component of forecast errors. Because analyst, firm, and time random effects are small, correctly accounting for them will not change the results drastically. This may explain why the present study's results are different from Keane and Runkle's, but are consistent with the findings of other previous studies.

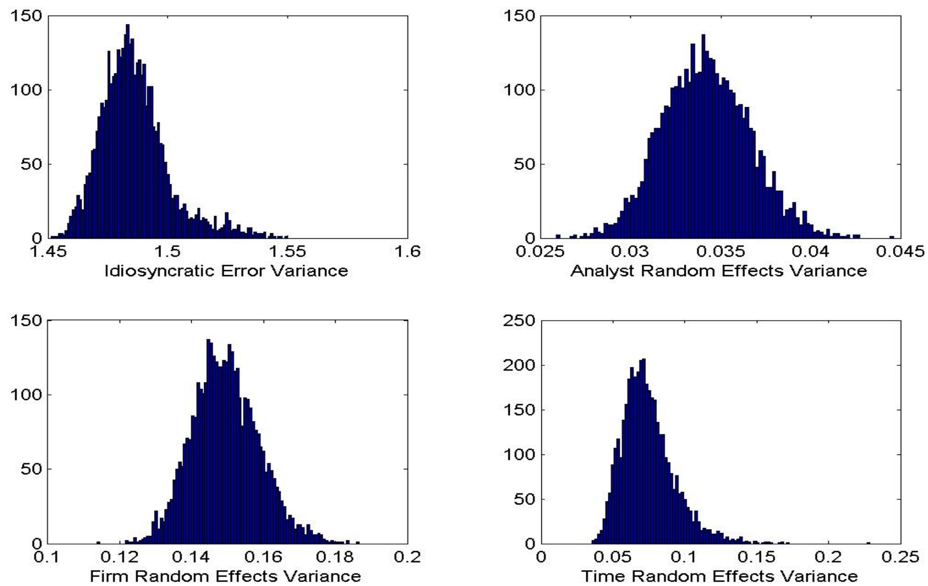


Figure 2.1 Histograms of Selected Parameters

The present study's results contrast the lone results obtained by Keane and Runkle (1998) which failed to reject the hypothesis of rationality. They stated "We fail to reject the hypothesis of rationality as long as we take into account two complications: (1) the correlation in a given period of analysts' forecast errors in predicting earnings for firms in the same industry and (2) discretionary asset write-downs, which affect earnings but are intentionally ignored by analysts when they make earnings forecasts." The above two complications alone may not be the reasons that they found rationality in analysts' forecasts. First, the treatment of forecast error correlation in Keane and Runkle (1998) is not complete and is based on assumptions that do not seem to be justified by the data. Second, the effects of special charges may not be large as shown before. Even after correctly accounting for forecast error correlations and dropping large forecast errors, the present study still finds significant evidence to reject the null hypothesis of unbiasedness.

Ruling out the above two reasons, the difference between Keane and Runkle (1998) and the present study as well as most of previous research may be explained by the much smaller sample used by the former authors. They only selected six four-digit Standard Industrial Classification industries to analyze. Each industry only has 3 to 5 firms and each industry only has 300 to 600 observations. They found that the forecasts of one industry they studied, the airline industry, are not rational. They attributed this to the aggregate shocks happened to the industry. In a related study that analyzes forecasts at the individual analyst level, Wu (2006) found that there is heterogeneity in the degrees of analysts' bias and inefficiency. The forecasts of some analysts, especially those of some firms that these analysts follow, can be regarded as unbiased and efficient. Therefore, it is quite possible to find subsets of firms and analysts for which the rationality condition can not be rejected.

2.5.2 Estimation results regarding efficiency

When estimating equations (2.6) to (2.8), the Gibbs sampler is run for 5000 iterations. The first 1000 iterations are discarded as the burn-in period. Different chains are run with different and over-dispersed starting values. The commonly used convergence tests (e.g., the Geweke diagnostic test and the Gelman and Rubin diagnostic test) are performed on the parameters $\alpha, \sigma^2, \sigma_a^2, \sigma_f^2, \sigma_t^2$, various β parameters, and randomly selected random effects. All of the parameters tested passed the convergence tests.

Estimation results of equations (2.6) to (2.8) are presented respectively in Panel A to Panel C in Table 2.2. From Table 2.2, the posterior standard deviations are very small compared with the posterior means, indicating that the parameters are quite accurately measured. The results regarding variances of random effects are similar to the previous findings about analysts' forecasting bias, with the exception that the firm random effect variances are significantly smaller than the variance corresponding to equation (2.4).

The posterior means and standard deviations of α in Table 2.2 are very close to those in Table 2.1 for the five samples. The posterior means of α in the ByFirm and ByAnalyst samples are not significantly different from zero, which indicates that forecasts are not biased in these two samples. However, the posterior means of β are significantly greater than zero, especially the coefficients of $PQFE$. This shows that although forecasts in the ByFirm and ByAnalyst samples are not biased, they are still inefficient with respect to the variables included in this study. Since the efficiency results for the ByEPSDev, ByFirm, ByAnalyst, and ByAnalystFirm samples are qualitative similar to the complete sample, the following efficiency analysis focuses on the complete sample.

Table 2.2 Estimation Results Regarding EPS Forecasts' Inefficiency

Panel A

	All	ByAnalystFirm	ByFirm	ByAnalyst	ByEPSDev
	Post. Mean	Post. Mean	Post. Mean	Post. Mean	Post. Mean
α	-0.090 (0.049)	-0.057 (0.043)	-0.056 (0.023)	-0.027 (0.034)	-0.107 (0.043)
$\beta_1\ (PQFE)$	0.330 (0.003)	0.307 (0.003)	0.326 (0.004)	0.309 (0.004)	0.340 (0.004)
$\beta_2\ (PQEC)$	0.019 (0.002)	0.011 (0.002)	0.010 (0.002)	0.007 (0.002)	0.036 (0.003)
σ^2	1.319 (0.005)	0.913 (0.004)	0.653 (0.003)	0.804 (0.005)	0.835 (0.005)
σ_a^2	0.012 (0.001)	0.008 (0.001)	0.009 (0.000)	0.011 (0.001)	0.014 (0.001)
σ_f^2	0.067 (0.004)	0.048 (0.003)	0.030 (0.003)	0.043 (0.003)	0.048 (0.004)
σ_t^2	0.059 (0.015)	0.049 (0.012)	0.043 (0.011)	0.047 (0.012)	0.052 (0.013)

Panel B

	All	ByAnalystFirm	ByFirm	ByAnalyst	ByEPSDev
	Post. Mean	Post. Mean	Post. Mean	Post. Mean	Post. Mean
α	-0.087 (0.036)	-0.040 (0.040)	-0.028 (0.029)	-0.045 (0.049)	-0.137 (0.036)
$\beta_1\ (PQFE)$	0.321 (0.003)	0.300 (0.003)	0.319 (0.004)	0.302 (0.004)	0.328 (0.004)
$\beta_2^R\ (PQEC_R)$	0.001 (0.002)	-0.003 (0.002)	-0.003 (0.002)	-0.006 (0.003)	0.002 (0.002)
$\beta_2^M\ (PQEC_M)$	0.065 (0.004)	0.048 (0.003)	0.048 (0.004)	0.044 (0.005)	0.099 (0.006)
σ^2	1.317 (0.005)	0.912 (0.004)	0.652 (0.003)	0.803 (0.005)	0.833 (0.006)
σ_a^2	0.012 (0.001)	0.008 (0.001)	0.009 (0.001)	0.011 (0.001)	0.014 (0.001)
σ_f^2	0.066 (0.005)	0.047 (0.003)	0.030 (0.003)	0.041 (0.004)	0.047 (0.004)
σ_t^2	0.059 (0.015)	0.049 (0.012)	0.044 (0.011)	0.048 (0.012)	0.052 (0.013)

Panel C

	All	ByAnalystFirm	ByFirm	ByAnalyst	ByEPSDev
	Post. Mean	Post. Mean	Post. Mean	Post. Mean	Post. Mean
α	-0.115 (0.030)	-0.083 (0.042)	-0.036 (0.035)	-0.044 (0.038)	-0.126 (0.044)
$\beta_1^P\ (PQFE_P)$	0.387 (0.007)	0.374 (0.006)	0.374 (0.007)	0.367 (0.008)	0.362 (0.009)
$\beta_1^N\ (PQFE_N)$	0.289 (0.004)	0.256 (0.005)	0.285 (0.005)	0.256 (0.006)	0.314 (0.005)
$\beta_2^{UR}(PQEC_{UR})$	-0.022 (0.004)	-0.024 (0.004)	-0.013 (0.004)	-0.023 (0.005)	-0.030 (0.006)
$\beta_2^{DR}(PQEC_{DR})$	0.022 (0.004)	0.016 (0.004)	0.006 (0.004)	0.016 (0.005)	0.029 (0.006)
$\beta_2^{UM}(PQEC_{UM})$	0.031 (0.005)	0.028 (0.005)	0.035 (0.005)	0.017 (0.006)	0.090 (0.009)
$\beta_2^{DM}(PQEC_{DM})$	0.103 (0.007)	0.065 (0.006)	0.058 (0.007)	0.081 (0.008)	0.100 (0.009)
σ^2	1.315 (0.005)	0.910 (0.004)	0.651 (0.003)	0.801 (0.005)	0.831 (0.005)
σ_a^2	0.012 (0.001)	0.008 (0.001)	0.009 (0.001)	0.011 (0.001)	0.014 (0.001)
σ_f^2	0.067 (0.004)	0.047 (0.003)	0.039 (0.003)	0.041 (0.003)	0.049 (0.004)
σ_t^2	0.058 (0.015)	0.048 (0.012)	0.044 (0.011)	0.047 (0.012)	0.052 (0.014)

Notes:

1. The regression specifications are as follows.

Panel A: $y_{aft+1} = \alpha + \beta_1 PQFE + \beta_2 PQEC + \mu_a + \mu_f + \mu_{t+1} + \varepsilon_{aft+1}$

Panel B: $y_{aft+1} = \alpha + \beta_1 PQFE + \beta_2^R PQEC_R + \beta_2^M PQEC_M + \mu_a + \mu_f + \mu_{t+1} + \varepsilon_{aft+1}$

Panel C:
$$y_{aft+1} = \alpha + \beta_1^P PQFE_P + \beta_1^N PQFE_N + \beta_2^{UR} PQEC_{UR} + \beta_2^{DR} PQEC_{DR} + \beta_2^{UM} PQEC_{UM} + \beta_2^{DM} PQEC_{DM} + \mu_a + \mu_f + \mu_{t+1} + \varepsilon_{aft+1}$$

2. Post. Mean refers to the mean of the parameter draws in the MCMC chain. Post. Std. Dev. refers to the standard deviation of the parameter draws in the MCMC chain.

The main focus of this section is the coefficients of various regressors and their implications for analysts' forecasting inefficiency. Panel A gives the results of equation (2.6), the baseline regression. It shows that the posterior mean of the coefficient of $PQFE$ is 0.330 and significantly greater than zero. Its sign means that there is a positive relationship between $CQFE$ and $PQFE$, i.e., if an analyst makes the mistake of over-estimating last quarter's EPS, he tends to over-estimate the current quarter's EPS, and vice versa. This relationship indicates that analysts tend to be slow in adjusting their forecasting practices to take into account their past errors. The previous literature (e.g., Ali, Klein and Rosenfeld (1992)) also documented such positive relationships. The coefficient of $PQEC$ is 0.019. This provides evidence that forecasts are inefficient with respect to past earnings changes. Previous studies using annual EPS measures estimate such coefficient at around 0.08. The difference may be explained by the shorter forecast horizon used in the present study or by the difference in the structure of the econometric model used.

Panel B gives the results of equation (2.7), which distinguishes $PQEC$ into earnings momentum and earnings reversion. The posterior mean of the coefficient on $PQEC_M$ is 0.065 and highly significant, whereas, the posterior mean of the coefficient on $PQEC_R$ is 0.001 and not significant. The above results contrast sharply with Constantinou, Forbes, and Skerratt (2003), who found β_2^M around 0.50 and β_2^R around -0.12. However, as explained before, their definitions of $PQEC_M$ and $PQEC_R$ contain information not available to analysts at the time of issuing their forecasts. Using only the information available to analysts, the present findings indicate that analysts under-react to the information contained in the $PQEC$ only in EPS momentum cases. This suggests that analysts consider momentum in earnings to be not sustainable and are very cautious to give a forecast of continuous trend.

Panel C gives the results of equation (2.8), which distinguishes between upward earnings momentum/reversion and downward earnings momentum/reversion, as well as different signs of $PQFE$. Significant positive serial correlations are found between forecast errors. The coefficient on $PQFE_P$ (0.387) is found to be larger than the coefficient on $PQFE_N$ (0.289). This suggests that analysts are quicker in adjusting their past mistakes when they over-estimated. The coefficient on downward momentum (0.103) is larger than the coefficient on upward momentum (0.031), which indicates analysts' general optimism and their reluctance to confirm downward EPS trends. The coefficient on downward reversion is 0.022 and the coefficient on upward reversion is -0.022. In the case of downward reversion ($PQEC$ - & $lagPQEC$ +), the EPS is expected to increase. In the case of upward reversion ($PQEC$ + & $lagPQEC$ -), the EPS is expected to decrease. This indicates that analysts over-react to the information contained in $PQEC_R$ as they overshoot the degree of earnings reversion.

2.5.3 Estimated time random effects

One advantage of the Bayesian estimation of the three-way random effects model proposed here is that all random effects are explicitly estimated. In previous research, the random effects are not of direct interest and they are treated as nuisance parameters. Since the magnitude of time-specific random effects could offer insights about analysts' forecasting behavior, the posterior means of the 31 time-specific random effects in equation (2.7) for the complete sample are shown in Figure 2.2 along with their 5th and 95th percentiles. Figure 2.2 shows that the estimated random effects for many quarters are fairly large, which indicates that analysts often failed to make good estimates of systematic shocks to the economy. For example, the largest negative time-specific random effect is -0.5, which happened during the 4th quarter of 2001. This is probably due to analysts' inability to estimate the negative impact of the events occurred during the 3rd quarter of 2001 (i.e., September 11 and its aftermath) and over-estimated too much the overall earnings levels for that quarter.

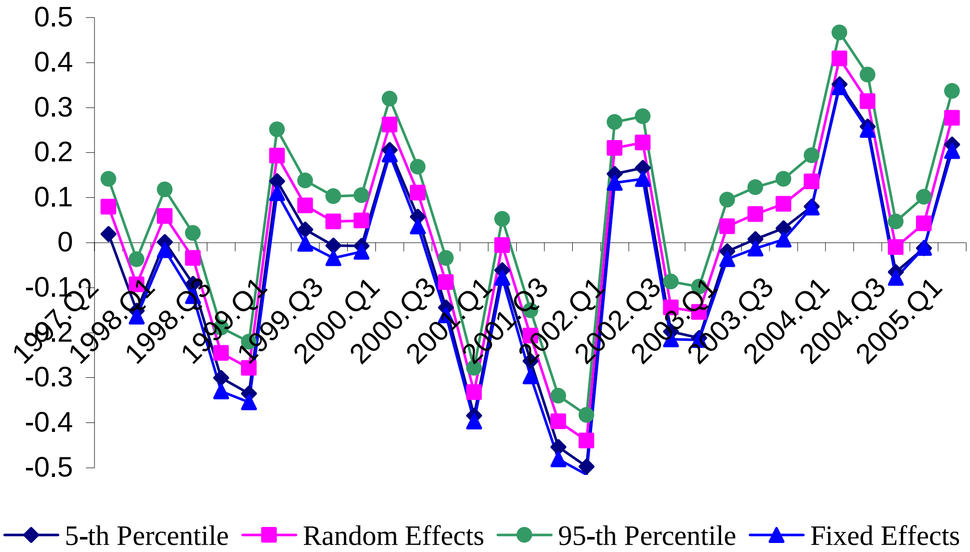


Figure 2.2 Estimated Time Random Effects

2.6 Conclusion

Using a three-way random effects model, the present study shows that consistent with most of the previous literature, as a population, analysts' forecasts are biased and inefficient. The results provide evidence that the three-way random effects model is appropriate to study analyst EPS forecast, as the variances of all three random effects are significantly larger than zero.

Analysts are found to be systematically optimistic. As for efficiency, the results show that if an analyst over-estimates EPS in one period, he will tend to over-estimate EPS in the next period, and vice versa. Analysts are found to be quicker in adjusting their past mistakes when they over-estimated than when they under-estimated. Analysts are found to under-react to previous quarter EPS changes in the cases of earnings momentum and over-react to previous quarter EPS changes in the cases of earnings reversions. The under-reaction is more severe in the cases of downward momentum, which indicates analysts' general optimism.

The present study offers additional evidence on the irrationality of analyst forecasts and reconciles previous contradicting results. It shows that accounting for the two reasons identified by Keane and Runkle (1998), namely forecast error correlations and special charges, is not sufficient to bring rationality to analyst forecasts. The present results suggest that the much smaller sample used by Keane and Runkle may be the reason of why their finding are different compared to most of the literature. The present study also points out an inconsistency in the construction of variables in Constantinou, Forbes and Skerratt (2003). Using correctly constructed variables, the present study reaches different conclusions.

The present study extends the research on stock analysts by correctly accounting for forecast error correlations. Previous studies either did not account for forecast error correlations, or imposed assumptions not supported by the data. The present study also demonstrates that analysts respond asymmetrically different types of past information.

2.7 References

Amir, Eli and Yoav Ganzach (1998), Over-reaction and under-reaction in analysts' forecasts. Journal of Economic Behavior and Organization, 37, 3, 333-347.

Ali, A., A. Klein, and J. Rosenfeld (1992), Analysts' use of information about permanent and transitory earnings components in forecasting annual EPS, Accounting Review, 67, 1, 183-198.

Abarbanell, J.S. (1991), Do analysts' earnings forecasts incorporate information in prior stock price changes? Journal of Accounting and Economics. 14, 2, 147-165.

Abarbanell, J. and V. Bernard (1992), Tests of analysts' over-reaction/under-reaction to earnings information as an explanation for anomalous stock price behavior. Journal of Finance, 47, 1181-1207.

Brown, Lawrence D., Robert L. Hagerman, Paul A. Griffin, and Mark E. Zmijewski (1987), Security analyst superiority relative to univariate time-series models in forecasting quarterly earnings. Journal of Accounting and Economics 9, 61-87.

Carlos Capistran Carmona, (2005), Bias in Federal Reserve inflation forecasts: is the Federal Reserve irrational or just cautious? Working paper.

Clement, M. (1999), Analyst forecast accuracy: do ability, resources, and portfolio complexity matter? Journal of Accounting and Economics, 27, 285-303.

Constantinou, Ph. Constantina, William P. Forbes, and Len Skerratt (2003), Analyst under-reaction in the United Kingdom. Financial Management, Summer, 93-106.

Crichfield, Timothy, Thomas R. Dyckman and Josef Lakonishok (1978), An evaluation of Security Analysts' Forecasts. Accounting Review, 53, 651-668.

De Bondt, Werner F. M., and Richard H. Thaler (1990), Do security analysts overreact? American Economic Review 80, 52-57.

Dische, A. and H. Zimmermann (1999), Consensus forecasts of corporate earnings changes and the performance of Swiss stocks. Journal of Investing Spring 19-26.

Easterwood, J. and S. Nutt (1999), Inefficiency in analysts' earnings forecasts: systematic misreaction or systematic optimism? Journal of Finance, 54, 5, 1777-1797.

Elliot, J., D. Philbrick, and C. Wiedman (1995), Evidence from archival data on the relation between security analysts' forecast errors and prior forecast revisions, Contemporary Accounting Research, 11, 919-938.

Goldstein, Harvey (2003), Multilevel statistical models. Kendall's library of statistics 3, Arnold, London.

Jacob, J., T. Lys, M. Neale (1999), Expertise in forecasting performance of security analysts. Journal of Accounting and Economics, 28, 51-82.

Keane, Michael P., and David E. Runkle (1998), Are financial analysts' forecast of corporate profits rational? Journal of Political Economy, 106, 4, 768-805.

Kothari, S.P. (2001), Capital markets research in accounting. Journal of Accounting and Economics 31, 105-232.

Lys, T. and S. Sohn (1990), The association between revisions of financial analysts' earnings forecasts and security-price changes, Journal of Accounting and Economics 13, 4, 341-363.

Mincer, J. and V. Zarnowitz (1969), The Evaluation of Economic Forecasts, in: J. Mincer (ed.) Economic Forecasts and Expectations. New York: National Bureau of Economic Research.

Mikhail, M., Beverly R. Walther; Richard H. Willis (1997), Do security analysts improve their performance with experience? Journal of Accounting Research, 35, 131-166.

O'Brien, Patricia C. (1988), Analysts' Forecasts as Earnings Expectations. Journal of Accounting and Economics, 10, 53-83.

Womack, Kent L. (1996), Do brokerage analysts recommendation have value? Journal of Finance 51, 137-167.

Wu, Jingtao (2007), A Bayesian Hierarchical Study of Analyst Bias and Inefficiency. Working paper.

2.8 Appendix

This appendix describes the complete Gibbs sampler with conditional posterior distributions for each set of parameters. Since the Gibbs sampling can be easily extended to include more independent variables to accommodate the estimation of equations (5) to (7), the notation in this section only refers to equation (4).

$$y_{aft+1} = \alpha + \mu_a + \mu_f + \mu_{t+1} + \varepsilon_{aft+1}$$

$$\alpha \sim N(\mu_\alpha, V_\alpha)$$

$$\varepsilon_{aft+1} \overset{iid}{\sim} N(0, \sigma^2)$$

$$\mu_a \overset{iid}{\sim} N(0,\sigma_a^2)\text{, for } a=1,\cdots,A$$
$$\mu_f \overset{iid}{\sim} N(0,\sigma_f^2)\text{, for } f=1,\cdots,F$$
$$\mu_{t+1} \overset{iid}{\sim} N(0,\sigma_t^2)\text{, for } t=1,\cdots,T$$
$$\sigma^2 \sim IG(a,b)$$
$$\sigma_a^2 \sim IG(a_a,b_a)$$
$$\sigma_f^2 \sim IG(a_f,b_f)$$
$$\sigma_t^2 \sim IG(a_t,b_t)$$

Each observation in the sample has its own μ_a, μ_f, μ_{t+1} and is normally distributed as.

$$y_{aft+1} \mid \alpha, \mu_a, \mu_f, \mu_{t+1} \sim N(\alpha+\mu_a+\mu_f+\mu_{t+1}, \sigma^2)\,.$$

The joint posterior distribution for parameters of the proposed model can be written as,

$$p(\Gamma \mid data) \propto \prod_{aft}^{N} \Phi(y_{aft+1} \mid \alpha, \mu_a, \mu_f, \mu_{t+1}, \sigma^2) \prod_{a}^{A} \Phi(\mu_a \mid \sigma_a^2) \prod_{f}^{F} \Phi(\mu_f \mid \sigma_f^2) \prod_{t}^{T} \Phi(\mu_{t+1} \mid \sigma_t^2)$$
$$*\Phi(\alpha \mid \mu_\alpha, V_\alpha)\, p(\sigma^2 \mid a,b)\, p(\sigma_a^2 \mid a_a,b_a)\, p(\sigma_f^2 \mid a_f,b_f)\, p(\sigma_t^2 \mid a_t,b_t)$$

Complete Posterior Conditional for α

$$p(\alpha \mid \Gamma_{-\alpha}, data) \propto \prod_{aft}^{N} \Phi(y_{aft+1} \mid \alpha, \mu_a, \mu_f, \mu_{t+1}, \sigma^2)\Phi(\alpha \mid \mu_\alpha, V_\alpha)$$

$$\alpha \mid \Gamma_{-\alpha}, data \sim N(D_\alpha d_\alpha, D_\alpha)$$
$$D_\alpha = (I_N{}' I_N / \sigma^2 + V_\alpha^{-1})^{-1}$$
$$d_\alpha = \sum_{N} (y_{aft+1} - \alpha - \mu_a - \mu_f - \mu_{t+1}) / \sigma^2 + V_\alpha^{-1}\mu_a$$

Complete Posterior Conditional for μ_a, $a=1,\cdots,A$

$$p(\mu_a \mid \Gamma_{-\mu_a}, data) \propto \prod_{ft}^{N} \Phi(y_{aft+1} \mid \alpha, \mu_a, \mu_f, \mu_{t+1}, \sigma^2)\Phi(\mu_a \mid \sigma_a^2)$$

$$\mu_a \mid \Gamma_{-\mu_a}, data \sim N(D_{\mu_a} d_{\mu_a}, D_{\mu_a})$$

$$D_{\mu_a} = (I_{N_a}{}' I_{N_a} / \sigma^2 + (\sigma_a^2)^{-1})^{-1}$$

$$d_{\mu_a} = \sum_{N_a} (y_{aft+1} - \alpha - \mu_f - \mu_{t+1}) / \sigma^2$$

Similarly for μ_f and μ_{t+1}.

Complete Posterior Conditional for σ^2

$$p(\sigma^2 \mid \Gamma_{-\sigma^2}, data) \propto \prod_{aft}^{N} \Phi(y_{aft+1} \mid \alpha, \mu_a, \mu_f, \mu_{t+1}, \sigma^2) p(\sigma^2 \mid a, b)$$

$$\sigma^2 \mid \Gamma_{-\sigma^2}, data \sim IG(a + N/2, (b^{-1} + \sum_{N} (y_{aft+1} - \mu_a - \mu_f - \mu_{t+1})^2 / 2)^{-1})$$

Complete Posterior Conditional for σ_a^2

$$p(\sigma_a^2 \mid \Gamma_{-\sigma_a^2}, data) \propto \prod_{a=1}^{A} \Phi(\mu_a \mid \sigma_a^2) p(\sigma_a^2 \mid a_a, b_a)$$

$$\sigma_a^2 \mid \Gamma_{-\sigma_a^2}, data \sim IG(a_a + N_a / 2, (b_a^{-1} + \sum_{N_a} (\mu_a)^2 / 2)^{-1})$$

Similarly for σ_f^2 and σ_t^2.

CHAPTER 3. A BAYESIAN HIERARCHICAL STUDY OF ANALYST BIAS AND INEFFICIENCY

3.1 Abstract

Using a Bayesian hierarchical model, the present study shows that, consistent with previous studies, analysts' forecasts are biased and inefficient as a population. The Bayesian hierarchical model allows us to avoid making scaling transformations to the original data. It also allows us to relax the unrealistic assumption of analyst homogeneity. It is shown that there is heterogeneity in the degrees of analysts' bias and inefficiency. The forecasts of some analysts, especially those of some firms that these analysts follow, can be regarded as unbiased and efficient. Based on the results of the hierarchical model, two approaches to forecast earnings surprises are proposed. The proposed measures are shown to be able to forecast earnings surprises with success rates that are statistically higher than 50%.

3.2 Introduction

Stock analysts who make forecasts of earnings per share (EPS) are an interesting group of economic agents to study for several reasons. First, researchers have found that earnings forecasts appear to have economic value for investors (e.g., Womack (1996)). Second, analysts' forecasts have often been found to outperform time series models, suggesting that analysts are rather good at what they do (e.g., Brown, Hagerman, Griffin, and Zmijewski (1987)). Third, the precision of analysts' forecasts represents an upper bound of the quality of earnings forecasts made by less sophisticated agents.

Although the importance of analysts is beyond dispute, the quality of their forecasts has become the subject of intense research and debate. An important category of research on analysts' forecasts examines their bias and inefficiency. In this literature, forecasts are usually considered to be biased if there is a systematic positive (or negative) difference between the forecasts and the actual EPS, whereas forecasts are typically labeled inefficient if they do not fully incorporate past information available at the time of issuing the forecasts. One of the most widely held beliefs in the literature is that analysts produce biased forecasts that are "too optimistic" (see, e.g., review by Kothari (2001)). In addition, numerous studies have documented analysts' inefficiency with respect to public information such as past earnings levels (e.g., De Bondt and Thaler (1990)), past earnings changes (e.g., Abarbanell and Bernard (1992)), extreme past earnings changes (e.g., Easterwood and Nutt (1999)), past returns (e.g., Lys and Sohn (1990), Abarbanell (1991), Ali, Klein and Rosenfeld (1992)), past forecast errors (e.g., Ali, Klein and Rosenfeld (1992)), and past forecast revisions (e.g., Amir and Ganzach (1998)), Mendenhall (1991)). Ali, Klein and Rosenfeld (1992), among others, document that analysts' forecasts are inefficient with respect to their most recent forecast error.

Another category of research focuses on the properties of individual analysts' forecasts, by studying the determinants of their accuracy (Mikhail, Walther and Willis (1997), Jacob, Lys and Neale (1999), Clement (1999)). These studies suggest that experience, the size of the brokerage firm that an analyst works for, and the number of firms and industries followed by an analyst affect forecast accuracy.

Because each firm is usually followed by several analysts, there are usually several EPS forecasts given by different analysts for a given firm and time. The common practice of the previous literature is to pool the consensus forecasts (the mean or median of all available forecasts of a given firm at a given time) across firms and across time, and then perform ordinary least squares (OLS) regression analysis on the pooled sample. This approach is likely to exhibits problems of cross-sectional dependence among observations, because different forecasts' qualities for a particular firm (or a particular analyst) tend to be correlated. Furthermore, since the previous literature aggregates information from all firms, it almost universally employs some sort of scaling procedure to the EPS and forecasts. The need for scaling arises because forecast errors may be heteroskedastic across firms (i.e., the deviations of forecasts from the actual EPS may depend on the levels of share prices or EPS). The most popular choices for scaling are previous share prices and EPS. The unintended consequence of scaling is that it could introduce unnecessary noises in the econometric system.

Because of the above potential shortcomings of the OLS method commonly employed in the previous literature, the present study looks at analysts' forecasts bias and inefficiency using a Bayesian hierarchical model. In the hierarchical model proposed here, the scaling procedure is not necessary. Most importantly, the hierarchical model does not treat analysts as homogenous as the previous literature implicitly did by pooling forecasts across time and firms. The reasons to treat analysts as heterogeneous are twofold. From a theoretical standpoint, it is reasonable to allow for analysts to have different abilities, leading to different degrees of bias and inefficiency in their forecasts. The previous studies have identified factors that affect the forecast accuracy, suggesting that analysts are likely to have different forecasting skills. From a practical standpoint, investors appear to think that analysts have different skills. Evidence in this regard is that the magazine *Institutional Investor* conducts annual polls of money managers regarding analysts forecast qualities. The top three vote getters in each industry are called All-Americans and are highly rewarded for this honor. In sum, it seems reasonable to allow for differential ability among analysts when setting up the econometric model.

The Bayesian hierarchical model proposed in the present study could yield bias and inefficiency estimates not only for a "representative" analyst like the previous literature, but also for each individual analyst as well as each individual analyst-firm combination. It allows us to avoid making scaling transformations to the original data. Such transformations are unavoidable in past research methods but may introduce problems of their own. In addition, the Bayesian hierarchical model allows us to relax the unrealistic assumption of analyst homogeneity. Using the three-level Bayesian hierarchical model, consistent with previous studies, it is found that as a population, analysts' forecasts are biased and inefficient. Analysts are systematically optimistic and their forecasts are too extreme. As for efficiency, the results show that if an analyst over-estimates EPS in one period, he will tend to over-estimate EPS next period, and vice versa. There is also evidence that if a firm's EPS in the current quarter is greater than last quarter's, analysts tend to under-estimate next quarter's EPS, and vice versa. The main contribution of the present study is the analysis of forecasts at the individual analyst level. Considerable heterogeneity in the degrees of analysts' bias and inefficiency is found. The forecasts of some analysts, especially those of some firms that these analysts follow, can be regarded as unbiased and efficient. Based on the results of the hierarchical model, two approaches to forecast earnings surprises are proposed here. The proposed measures are shown to be able to forecast earnings surprises with success rates that are statistically higher than 50%.

The rest of the paper is organized as follows. Section 3 introduces the Bayesian hierarchical model. Section 4 lays out the estimation framework. Section 5 and 6 discuss results regarding bias and inefficiency respectively. Section 7 presents the results on earnings surprises forecasting. Section 8 concludes.

3.3 The Bayesian Hierarchical Model

The study of analysts has a hierarchical, nested, or clustered structure, i.e., firms are grouped within their corresponding analysts who follow them. The basic variation is therefore at two levels, between analyst-firms and between analysts. Analyst-firm refers to the unique combination of one analyst and one firm he follows, since each firm can be followed by multiple analysts and each analyst can follow multiple firms. As shown later, Analyst-firms from the same analysts are found to be more alike in their EPS forecast qualities than analyst-firms chosen at random. One possible explanation for this difference is that analyst-firms from the same analyst have similar forecasting difficulty levels due to a common business environment, as analysts tend to follow companies in the same sector. Another possible explanation is that analysts are heterogeneous in their forecasting skills, with some of them performing consistently better than others.

The longitudinal structure of the data, with multiple forecasts for each analyst-firm, introduces an additional level, namely a between-forecast-within-analyst-firm (BfWaf) level. The BfWaf level consists of each analyst-firm's time series observations of forecasts and EPS. Therefore, the BfWaf level is level 1, analyst-firms are level 2, and analysts are level 3. The existence of the above data hierarchies is neither accidental nor ignorable. Once a group (e.g., several analyst-firms form an analyst group) is established, often it tends to become differentiated from the other groups. This differentiation implies that the group and its members both influence and are influenced by the group membership. To ignore this relationship risks overlooking the importance of group effects, and may also render invalid many of the traditional statistical analysis techniques used for studying data relationships (Goldstein (2003)). This is true because if the qualities of one analyst's forecasts for different quarters of a particular firm tend to be similar, they would provide less information than if the same number of forecasts were for different firms. Put another way, when looking only at levels 1 and 2, the basic unit for purpose of comparison should be the analyst-firm, not the forecast. The function of the forecasts can be seen as providing, for each analyst-firm, an estimate of that analyst-firm's quality. Increasing the number of forecasts per analyst-firm would increase the precision of those estimates but not change the number of analyst-firms being compared. Beyond a certain point, simply increasing the numbers of forecasts hardly improves things at all. However, increasing the number of analyst-firms to be compared, with the same or somewhat smaller number of forecasts per analyst-firm, considerably improves the precision of the comparisons. Analogously, the same relationship exists between levels 2 and 3.

One important issue arises when one wishes primarily to have information about each analyst-firm in a sample. In this instance, because of the large number of analyst-firms, estimation by OLS involves a large number of parameters which is likely to result in overfitting. To further complicate matters, some analyst-firms may have rather few quarterly observations, in which case OLS is likely to yield imprecise estimates. In such situations, because the hierarchical structure regards analyst-firms as members of a population (analyst), population estimates of the mean and between-(analyst-firm) variations are used to obtain more precise estimates for each analyst-firm. Analogously, the same effects apply to level 3 (analyst) estimates.

Estimation of the advocated hierarchical structure can be implemented by maximum likelihood (ML) using iterative generalized least squares. However, the random effects are not directly estimated in ML, but are summarized into their estimated variances and covariances. If one is only interested in fixed effects, which are estimated directly, then ML will suffice. However, if one is interested in the specific random effects at levels 1 and 2, then an alternative estimation method is needed. The alternative estimation method proposed in the present paper is Bayesian estimation using Markov Chain Monte Carlo (MCMC). The Bayesian hierarchical model incorporates prior distribution assumptions and, based on successively sampling from conditional posterior distributions of the model parameters, yields chains of parameter values which are then used for making inferences. The advantage of the Bayesian hierarchical model is that it yields parameter estimates at each level. Therefore, one can get much more detailed information about cross-unit variations in addition of variances and covariances. The relevance of such information will become evident in later sections.

3.4 Estimation Framework

3.4.1 Testing for bias and inefficiency

Suppose at time t, analyst a follows firm f. The realization of firm f's next quarter EPS, E_{ft+1}, is unknown to analyst a. However, analyst a may obtain information about E_{ft+1} which allows him to issue a conditional forecast of E_{ft+1}, F_{aft}^{t+1}. If analyst a had perfect foresight, then $E_{ft+1} = F_{aft}^{t+1}$. Therefore, the following equation should hold for each t,

$$E_{ft+1} - E_{ft} = F_{aft}^{t+1} - E_{ft}\,. \tag{3.1}$$

In the absence of perfect foresight, equation (3.1) does not hold exactly. The following regression could be run to test the validity of (3.1):

$$E_{ft+1} - E_{ft} = \alpha_{af} + \beta_{af}(F_{aft}^{t+1} - E_{ft}) + \varepsilon_{aft},\ \ \varepsilon_{aft} \sim N(0, \sigma_{af}^2)\,. \tag{3.2}$$

The joint null hypothesis of $H0: \alpha_{af} = 0, \beta_{af} = 0$ is regarded as a test of whether analyst a's EPS forecasts for firm f are unbiased. If the null hypothesis is supported by the data, then the forecasts are regarded as unbiased.

If $(F_{aft}^{t+1} - E_{ft})$ is subtracted from both sides of equation (3.2), the transformed equation is,

$$E_{ft+1} - F_{aft}^{t+1} = \alpha_{af} + \beta_{af}^{*}(F_{aft}^{t+1} - E_{ft}) + \varepsilon_{aft},\ \ \varepsilon_{aft} \sim N(0, \sigma_{af}^2)\,. \tag{3.3}$$

The dependent variable of equation (3.3) is analyst a's forecast error of E_{ft+1}. Let I_t be a set of variables known as of time t, analyst a's forecasts are I_t-efficient if they comprise all available information in I_t. Therefore, the coefficients of the variables in I_t will be insignificant in regression (3.3) if forecasts are I_t-efficient. Two obvious candidates for inclusion in I_t are $E_{ft} - F_{aft-1}^{t}$, the forecast error of E_{ft}, and $E_{ft} - E_{ft-1}$, the difference of EPS from $t-1$ to t. Therefore, the extended version of equation (3.3) is,

$$E_{ft+1} - F_{aft}^{t+1} = \alpha_{af} + \gamma_{af}(F_{aft}^{t+1} - E_{ft}) + \delta_{af}(E_{aft} - F_{aft-1}^{t}) + \phi_{af}(E_{ft} - E_{ft-1}) + \tilde{\varepsilon}_{aft}, \tilde{\varepsilon}_{aft} \sim N(0, \tilde{\sigma}_{af}^2) . (3.5)$$

3.4.2 Data

The present study focuses on analysts covering the U.S. equity market. The following procedure is used to construct the sample used for estimation. First, analysts who give forecasts during 2004 are identified. Second, each identified analyst's forecast histories of the firms he follows are collected from historical public records. The EPS measures used are quarterly.[5] $E_{ft+1} - E_{ft}$ and $F_{aft}^{t+1} - E_{ft}$ are constructed by taking the differences between adjacent quarters for each analyst-firm.[6] Third, any analyst-firm combination that has observations fewer than 12 quarters is deleted from the sample. The final sample has 1428 analysts covering 3829 unique firms, with 14478 unique analyst-firm combinations. On average, analysts cover 16 firms, and the maximum number of firms covered by a single analyst is 51.

3.4.3 The Gibbs sampler

This section outlines the Gibbs sampler used to estimate the parameters in equation (3.2). The Gibbs sampler can be easily extended to include more independent variables to accommodate the estimation of equation (3.4). Without loss of generality, the notation in this section and the Appendix only refers to equation (3.2).

To simplify notation, equation (3.2) is rewritten into the following matrix form. Let N_{af} denote the number of observations for analyst a's forecast history for firm f. Stack the

---

[5] The above quarters are calendar quarters. For example, if a firm announces its earnings during the first three months of 2004, that earnings report will be classified as corresponding to the first quarter of 2004, regardless of the actual fiscal quarter of that firm.

[6] It is common practice for analysts to revise their forecasts before the actual earnings are announced. In this study, only analysts' most recent forecasts are kept in the sample.

N_{af} observations of changes in actual EPS, $E_{ft+1} - E_{ft}$, into a column vector y_f and the corresponding forecast errors, $F_{aft}^{t+1} - E_{ft}$, into a column vector x_{af}. Let X_{af} denote $[I_{N_{af}\times 1} \quad x_{af}]$ where $I_{N_{af}\times 1}$ is an N_{af} vector of ones, and define θ_{af} as the vector $[\alpha_{af} \quad \beta_{af}]'$. Then (3.2) can be rewritten as,

$$y_f = X_{af} \bullet \theta_{af} + \varepsilon_{af},\ \varepsilon_{af} \sim N(0, \sigma_{af}^2 I_{N_{af}}). \tag{3.5}$$

Priors are represented by:

$$\theta_{af} \overset{iid}{\sim} N(\theta_a, \Omega_a),\ f = 1, 2, \cdots, F_a;\ a = 1, 2, \cdots, A,$$
$$\sigma_{af}^2 \overset{iid}{\sim} IG(\bar{a}, \bar{b}),\ f = 1, 2, \cdots, F_a;\ a = 1, 2, \cdots, A,$$
$$\theta_a \overset{iid}{\sim} N(\theta_0, \Sigma),\ a = 1, 2, \cdots, A,$$
$$\Omega_a^{-1} \overset{iid}{\sim} Wishart([\rho_\Omega R_\Omega]^{-1}, \rho_\Omega),\ \ a = 1, 2, \cdots, A.,$$
$$\theta_0 \sim N(\eta, C),$$
$$\Sigma^{-1} \sim Wishart([\rho_\Sigma R_\Sigma]^{-1}, \rho_\Sigma),$$

where A denotes the number of analysts in the sample, and F_a denotes the number of firms followed by analyst a.

In the above framework, θ_0 is a population parameter and is given a multivariate normal prior with mean vector η and covariance matrix C. The θ_a vectors are analyst-level parameters and are given *iid* multivariate normal priors with mean vector θ_0 and covariance matrix Σ across analysts. The covariance matrix Σ measures the degree of heterogeneity among analysts and is given an inverse-Wishart prior with parameters ρ_Σ and R_Σ. The θ_{af} vectors are analyst-firm level parameters and are given *iid* multivariate normal priors with mean vector θ_a and covariance matrix Ω_a across analyst-firms. The covariance matrix Ω_a measures the degree of heterogeneity among analyst-firms within a given analyst, and is given an inverse-Wishart prior with parameters ρ_Ω and R_Ω. The σ_{af}^2 is the error variance and is given an inverse Gamma prior with parameters $\bar{a}$ and $\bar{b}$ by the properties of inverse Gamma distribution.

Assuming conditional independence across analysts, the joint posterior distribution for all of the parameters of the proposed model can be written as,

$$p(\Gamma \mid y) \propto \left\{ \prod_{a=1}^{A} \left[\prod_{f=1}^{F_a} p(y_f \mid X_{af}, \theta_{af}, \sigma_{af}^2) \, p(\theta_{af} \mid \theta_a, \Omega_a) \, p(\sigma_{af}^2 \mid a, b) \right] p(\theta_a \mid \theta_0, \Sigma) \, p(\Omega_a \mid \rho_\Omega, R_\Omega) \right\}$$
$$* \, p(\theta_0 \mid \eta, C) P(\Sigma \mid \rho_\Sigma, R_\Sigma)$$

where $\Gamma \equiv \left[\left\{ \theta_{af}, \sigma_{af}^2 \right\}, \left\{ \theta_a, \Omega_a \right\}, \theta_0, \Sigma \right]$ denotes all parameters of the model. (In the Appendix, Γ_{-x} denotes all parameters other than x.) Parameters in the above three-level hierarchical model can be divided into sets, namely, population, analyst, and analyst-firm. The population parameters affect all analysts, and consist of θ_0 and Σ. Analyst-specific parameters consist of θ_a and Ω_a, and analyst-firm specific parameters consist of θ_{af} and σ_{af}^2. Using a Gibbs sampler, the joint posterior of all these parameters can be analyzed one set at a time. By cycling repeatedly through draws of each parameter conditional on the remaining parameters, the Gibbs sampler produces a Markov chain of parameter draws whose joint distribution converges to the posterior. The conditional posterior distributions for each set of parameters and the Gibbs sampler are given in the Appendix. The notation and procedures of the Gibbs sampler closely follow those of Koop, Poirier, and Tobias (2006).

3.4.4 Prior choices and sensitivity analysis

The matrix Ω_a measures the variation among analyst-firms for each analyst, whereas the matrix Σ measures the variation across analysts. The inverse of Ω_a and Σ are given Wishart prior distributions with the following parameters,

$$\rho_\Omega = \rho_\Sigma = 15, \; R_\Sigma = \begin{bmatrix} 0.1^2 & 0 \\ 0 & 0.5^2 \end{bmatrix}, \; R_\Omega = \begin{bmatrix} 0.05^2 & 0 \\ 0 & 0.25^2 \end{bmatrix}.$$

The priors are chosen so that θ_{af} and θ_a are quite diffuse around their means and there is more analyst heterogeneity than there is analyst-firm heterogeneity within each analyst. As stated in the introduction, analysts are believed to have ability, therefore, the value of η and C are given as,

$$\eta = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, C = \begin{bmatrix} 0.05^2 & 0 \\ 0 & 0.5^2 \end{bmatrix}.$$

Finally, for the error variance σ^2_{af} , hyper-parameters are chosen as $\bar{a} = 3$ and $\bar{b} = 0.005$ so that the prior mean and prior standard deviation are equal to 0.005.

Similarly, for the estimation of equation (3.4), the priors are given as,

$$\rho_\Omega = \rho_\Sigma = 15,\ R_\Sigma = \begin{bmatrix} 0.1^2 & 0 & 0 & 0 \\ 0 & 0.5^2 & 0 & 0 \\ 0 & 0 & 0.1^2 & 0 \\ 0 & 0 & 0 & 0.1^2 \end{bmatrix},\ R_\Omega = \begin{bmatrix} 0.05^2 & 0 & 0 & 0 \\ 0 & 0.25^2 & 0 & 0 \\ 0 & 0 & 0.05^2 & 0 \\ 0 & 0 & 0 & 0.05^2 \end{bmatrix},$$

$$\eta = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix},\ \mathrm{C} = \begin{bmatrix} 0.05^2 & 0 & 0 & 0 \\ 0 & 0.25^2 & 0 & 0 \\ 0 & 0 & 0.05^2 & 0 \\ 0 & 0 & 0 & 0.05^2 \end{bmatrix}.$$

For the error variance σ^2_{af} , hyper-parameters are still chosen as $\bar{a} = 3$ and $\bar{b} = 0.005$.

To perform prior sensitivity analysis, inverse Wishart and inverse Gamma priors more and less diffuse than the ones presented above are used. The results indicate that the population level parameters are very robust to prior choices, the analyst level parameters are less robust (some analysts' results are robust, whereas some analysts' results are sensitive), and the analyst-firm level parameters are sensitive to prior choices. The reason for the sensitivity results is that the population (the whole sample) and some analysts (the third level) have many data points; therefore, the data dominate the priors. The analyst-firms (the second level) and some analysts do not have many data points. In these cases, the priors dominate the data and have an important effect on the estimation results.

3.5 Estimation Results Regarding Bias

The Gibbs sampler is run for 1100 iterations, discarding the first 100 of them as the burn-in period. Different chains were run with different and over-dispersed starting values. The commonly used convergence tests (e.g., the Geweke diagnostic test and the Gelman and Rubin diagnostic test) are performed on the population-level parameters (θ_0 and Σ) and randomly selected parameters from the analyst-level (θ_a and Ω_a) and the analyst-firm level (θ_{af} and σ^2_{af}). All of the parameters tested passed the convergence tests. The trace plots show that the simulated draws from different chains appear to settle down and explore the same region very quickly. In fact, after the first ten iterations or so, the progression of chains seems very similar. These provide suggestive evidence that the choice of the number of iterations and burn-in periods are adequate for the present application.

The estimated population parameters are reported in Table 3.1. From Table 3.1, it can be concluded that the posterior standard deviations are very small compared with the posterior means, indicating that the parameters are quite accurately measured. The estimate of intercept α_0 is -0.003 and significantly negative which is usually regarded in the literature as forecasts being too optimistic. The magnitude of the intercept estimate is considerably smaller than similar estimate in De Bondt and Thaler (1990), which report a value of -0.094. Since it is documented that analysts' bias gets larger with the increase of forecast horizon, the above difference is consistent with the fact that the estimation in the present study is based on quarterly data, whereas De Bondt and Thaler (1990) is based on annual data. The slope estimate β_0 is 0.664 and significantly less than one. The magnitude of the slope estimate is very close to De Bondt and Thaler (1990)'s estimate of 0.648. This presents evidence supporting the hypothesis that forecasts are too extreme. Ignoring the constant term in equation (3.2), a forecasted change of $1 in EPS is followed on average by an actual change of only 66.43 cents. Furthermore, when compared with Bondt and Thaler (1990)'s estimate based on annual EPS forecasts, analysts' forecasts do not seem to improve as the forecast horizon get shorter.

As for the covariance matrix results, there is evidence of heterogeneity across analysts, as both Σ_{11} and Σ_{22} are reasonably large. This heterogeneity shows that it may be rewarding to study each individual analyst rather than to simply draw inferences from the representative analyst of the aggregated sample. The covariance between the intercept and slope, Σ_{12}, implies a small and positive correlation coefficient of 0.03.

Table 3.1 Population Parameter Estimates Regarding Bias

	Post. Mean	Post. Std. Dev
α_0	-0.003	0.001
β_0	0.664	0.006
Σ_{11}	0.0003	1.70E-05
Σ_{12}	0.0001	0.0001
Σ_{22}	0.041	0.002

As stated in the introduction, one strong advantage of the Bayesian hierarchical model is that it can be used to obtain α_a and β_a estimates for each analyst, and α_{af} and β_{af} estimates for each analyst-firm. This allows us to study the degree of bias of each analyst and each analyst-firm. Because the parameters at the analyst and analyst-firm levels are too many to be presented individually, only kernel densities of posterior means $E(\alpha_a \mid y)$, $E(\beta_a \mid y)$, $E(\alpha_{af} \mid y)$, $E(\beta_{af} \mid y)$, and $E(\sigma_{af}^2 \mid y)$ are presented in Figure 3.1.[7] From the $E(\alpha_a \mid y)$ and $E(\alpha_{af} \mid y)$ density plots, the analyst and analyst-firm intercepts seem to be normally distributed and centered around zero. From the $E(\beta_a \mid y)$ and $E(\beta_{af} \mid y)$ density plots, the majority of analyst and analyst-firm slopes are less than one. These results provide evidence

[7] There are altogether (2+4)* $N + \sum_{i=1}^{N}(2+1)*F_i$ = (2+4)*1428+(2+1)*14478=52002 parameters in the second and third levels.

that analysts' forecasts are biased not only at the population level, but also at the analyst and analyst-firm levels. The results also provide strong evidence that there is substantial heterogeneity among analysts and among analyst-firms. Moreover, the forecasts of some analysts and analyst-firms could be considered as unbiased.

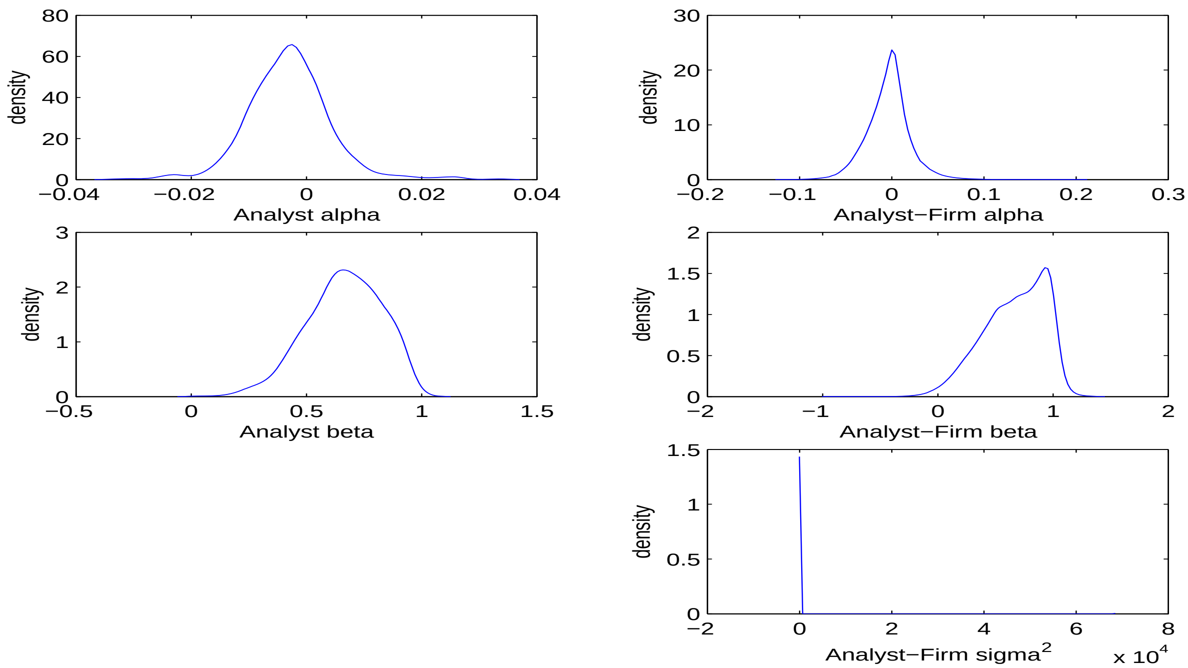


Figure 3.1 Densities of Posterior Means of Parameters Regarding Bias

One could go one step further beyond the simple visual interpretation of Figure 3.1 and calculate the Bayes factor in favor of the hypothesis that $\theta_{af} = [0 \quad 1]'$, which is given by,

$$B_{12}^{af} = \frac{\text{marginal posterior density of } \theta_{af} \text{ evaluated at } \theta_{af} = [0 \quad 1]'}{\text{marginal prior density of } \theta_{af} \text{ evaluated at } \theta_{af} = [0 \quad 1]'}$$
$$= \frac{p(\theta_{af} = [0 \quad 1]' \mid y)}{p(\theta_{af} = [0 \quad 1]')}$$

Within the hierarchical framework, the above test can also be carried out for individual analysts using the joint null hypothesis of $\alpha_a = 0$ and $\beta_a = 1$, and for the analyst population level using the joint null hypothesis $\alpha_a = 0$ and $\beta_a = 1$, In terms of the Bayes factors, $B_{12}^{a} = p(\theta_a = [0 \quad 1]^{'} \mid y) / p(\theta_a = [0 \quad 1]^{'})$ and $B_{12}^{0} = p(\theta_0 = [0 \quad 1]^{'} \mid y) / p(\theta_0 = [0 \quad 1]^{'})$ could be calculated to evaluate the bias at analyst level and population level. Given the results in Table 3.1, it is not surprising that the calculated $B_{12}^{0} \approx 0$. The percentiles of B_{12}^{a} and B_{12}^{af} are presented in Table 3.2. It is evident from Table 3.2 that none of the Bayes factors favor the null hypothesis at the 5% percentile. However, some large values of B_{12}^{a} and B_{12}^{af} do suggest that data favor the null hypothesis. For example, the 50% percentile of B_{12}^{af} is 1481.75, which means data favor the null hypothesis by a factor of 1481.75 to 1. The above result could also be seen straightforwardly from Figure 3.1. In Figure 3.1, since most of the masses of B_{12}^{a} and B_{12}^{af} are less than 1, it is hard to imagine that the unbiasedness could be substantiated for all analysts and analyst-firms. However, given that β_a and β_{af} do have masses at 1, and α_a and α_{af} do have masses at 0, it would not be surprising that the null hypothesis can not be rejected for some analysts and analyst-firms. In sum, as a population, analysts' forecasts can not be considered as unbiased. However, the forecasts of some analysts and analyst-firms can be deemed as unbiased. Furthermore, it is straightforward to quantify the degree of biasedness through α and β estimates and/or the Bayes factors.

Table 3.2 Percentiles of the Bayes Factors

	B_{12}^{a}	B_{12}^{af}
5%	1.9E-05	0.014
10%	0.004	0.360
15%	0.055	2.601
50%	92.427	1481.75
85%	2299.6	25962
90%	3481.7	40867
95%	5948	75763

Note: The null hypothesis is $\theta_{af} = \left[0 \quad 1\right]^{'}$ and $\theta_a = \left[0 \quad 1\right]^{'}$. If the Bayes factor is greater than one, data support the null, and vice versa.

3.6 Estimation Results Regarding Efficiency

Estimation results of equation (3.4) are presented in Table 3.3. The prior sensitivity analysis and convergence tests results are very similar to those of equation (3.2); therefore they are not discussed here to avoid redundancy. From Table 3.3, the posterior standard deviations are very small compared with the posterior means, indicating that the parameters are quite accurately measured. The posterior mean of α_0 is -0.003 and significantly smaller than zero, the posterior mean of γ_0 is -0.2917 and significantly smaller than zero. This is expected as $\gamma_0 = \beta_0 - 1$ if the other two regressors are ignored.

The coefficient of $E_{ft} - F_{aft-1}^{t}$, δ_0, is 0.145 and significantly greater than zero. Its sign means that there is a positive relationship between $E_{ft} - F_{aft-1}^{t}$ and $E_{ft+1} - F_{aft}^{t+1}$, i.e., if analyst a makes the mistake of over-estimating E_{ft} by one dollar, he tends to over-estimate E_{ft+1} by 14.5 cents, and vice versa. This relationship indicates that analysts tend to be slow in adjusting their forecasting practices to take into account their past errors. Previous literature (e.g., Ali, Klein and Rosenfeld (1992)) also documented such positive relationship but with larger magnitudes. One contribution of the present study is that the interpretation of δ_{af} is more compelling from a behavioral point of view, since δ_{af} is derived for each analyst-firm. Both $E_{aft} - F_{aft-1}^{t}$ and $E_{aft+1} - F_{aft}^{t+1}$ are mistakes made by the same analyst for the same firm. If consensus earnings are used, the comparison base most likely will not be the same. This is because not all of the same analysts will give forecasts for a given firm year after year.

The coefficient of past changes in EPS, $E_{ft} - E_{ft-1}$ is 0.022 and significantly different form zero. Previous studies (e.g., Constantinou, Forbes, and Skerratt (2003)) usually find estimates similar to ϕ_0 around 0.08. The difference may be explained by the shorter forecast horizon used in the present study as explained before. This result provides evidence that forecasts are inefficient with respect to past earnings changes. Comparing with the results on past forecast error, the coefficient on past EPS change is relatively small.

Table 3.3 Population Level Parameter Estimates Regarding Efficiency

	Post. Mean	Post. Std. Dev		Post. Mean	Post. Std. Dev
α_0	-0.003	0.001	Σ_{44}	0.006	0.000
γ_0	-0.292	0.006	Σ_{12}	0.000	0.000
δ_0	0.145	0.004	Σ_{13}	0.000	0.000
ϕ_0	0.022	0.003	Σ_{14}	0.000	0.000
Σ_{11}	0.000	1.52E-05	Σ_{23}	0.012	0.001
Σ_{22}	0.038	0.002	Σ_{24}	-0.008	0.001
Σ_{33}	0.011	0.001	Σ_{34}	-0.003	0.000

As discussed in the previous section, one could also study the degree of inefficiency of each analyst and each analyst-firm. Because the parameters at the analyst and analyst-firm levels are too many to be presented individually, only kernel densities of posterior means $E(\alpha_a \mid y)$, $E(\gamma_a \mid y)$, $E(\delta_a \mid y)$, $E(\phi_a \mid y)$, $E(\alpha_{af} \mid y)$, $E(\gamma_{af} \mid y)$, $E(\delta_{af} \mid y)$, $E(\phi_{af} \mid y)$ are presented in Figure 3.2. In Figure 3.2, the majority of the masses of $E(\delta_a \mid y)$ and $E(\delta_{af} \mid y)$ are greater than zero. However, $E(\delta_a \mid y)$ and $E(\delta_{af} \mid y)$ have considerable amount of masses at zero, which means that there is no relationship between past forecast errors and future forecast errors for some analysts and analyst-firms. The distributions of $E(\phi_a \mid y)$ and $E(\phi_{af} \mid y)$ are centered around zero, which implies that for many analysts, past earnings changes could not explain their forecast errors.

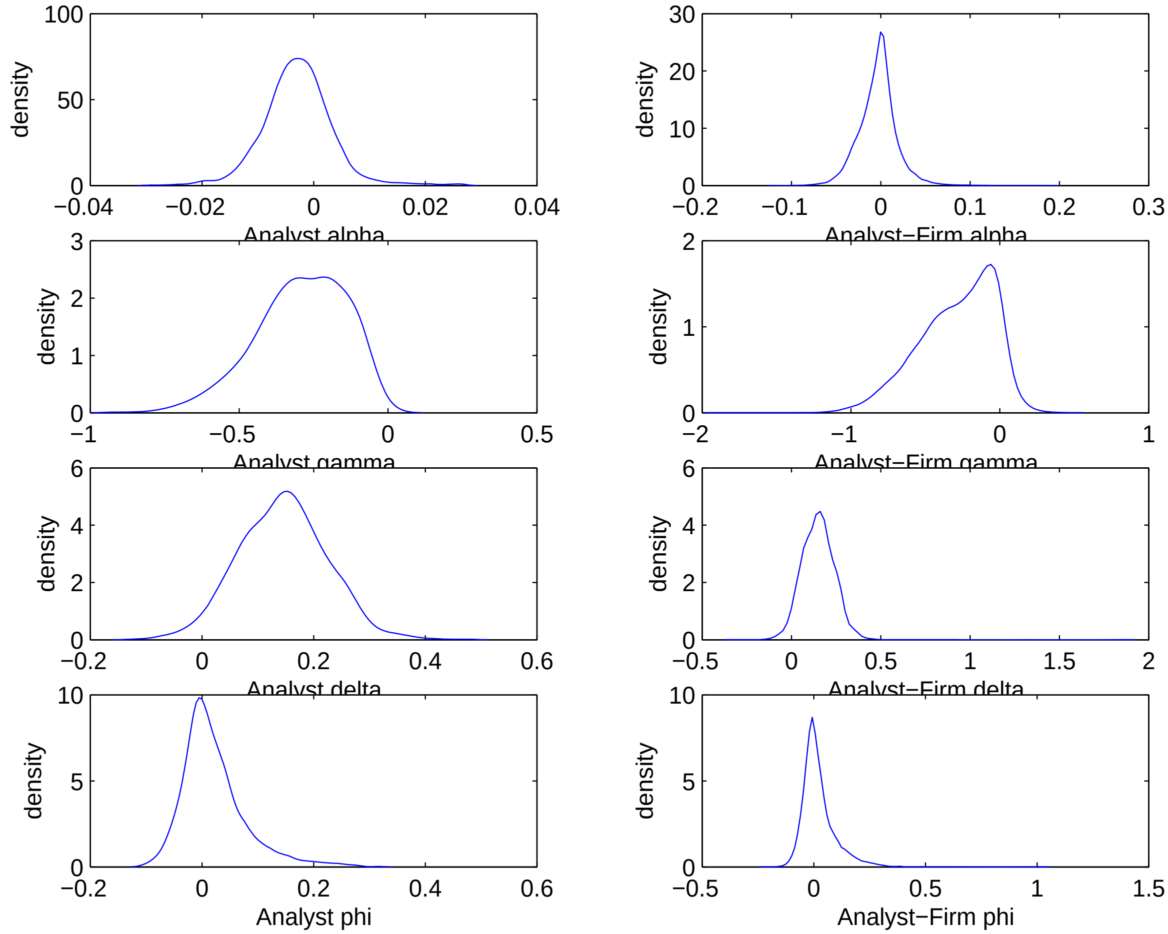


Figure 3.2 Densities of Posterior Means of Parameters Regarding Efficiency

3.7 Earnings Surprises Forecasting

Because the joint null hypothesis of $\theta_0 = \begin{bmatrix} 0 & 1 \end{bmatrix}'$ is not supported by the data, then equation (2) suggests that analysts' forecasts should be treated as signals of E_{ft+1} rather than as the conditional expectations of E_{ft+1}. Given data and Γ, after adjusting for forecast bias, in principle, a better adjusted forecast F_{aft}^{t+1*} can be calculated from analyst a's signal as,

$$F_{aft}^{t+1*} = E(E_{ft+1} \mid y) = E(\alpha_{af} \mid y) + E(\beta_{af} \mid y) F_{aft}^{t+1} + (1 - E(\beta_{af} \mid y)) E_{ft} .$$

The variance of the adjusted forecast F_{aft}^{t+1*} is $E(\sigma_{af}^{2} \mid y)$. When aggregating different analysts' forecasts for the same company, the *weighted mean* could be calculated by the following formula,

$$(Weighted\ \ Mean)_{ft+1} = \frac{\sum_{a}^{N_a} \frac{F_{aft}^{t+1*}}{E(\sigma_{af}^{2} \mid y)}}{\sum_{a}^{N_a} \frac{1}{E(\sigma_{af}^{2} \mid y)}}.$$

where N_a is the number of analysts who give forecasts for firm f at time t. In the above formula, the weight of each adjusted forecast is the inverse of $E(\sigma_{af}^{2} \mid y)$. Forecasts of better analysts (with smaller $E(\sigma_{af}^{2} \mid y)$) will carry bigger weights in the weighted mean.[8]

Investors usually care more about the consensus forecast rather than individual forecasts. The standard industry practice is to calculate the consensus forecast as the median or mean of all available analysts' forecasts. To forecast earnings surprises, the simple mean of the original forecasts (*Mean*) is compared with the weighted mean of the adjusted forecasts (*Weighted Mean*). If *Weighted Mean* > *Mean*, then *Weighted Mean* predicts that the consensus *Mean* will under-estimate the actual EPS. If the actual EPS is larger than *Mean*, then *Weighted Mean* is successful at predicting the positive earnings surprise. Similarly, if *Weighted Mean* < *Mean*, then *Weighted Mean* predicts that the consensus *Mean* will over-estimate the actual EPS. If the actual EPS is smaller than *Mean*, then *Weighted Mean* is successful at predicting the negative earnings surprise. Analogously, *Adjusted Median* can be calculated as the median of the adjusted forecasts. Predictions about earnings surprises can be derived from the comparison of *Adjusted Median* and *Median* (the median of original forecasts).

[8] Other weighting schemes, such as using the log of the Bayes factor of each analyst, yield results similar to the analysis in the present subsection.

To study the earnings surprises forecasting ability of *Weighted Mean* (*Adjusted Median*), let z_i denote the forecasting outcome of each *Weighted Mean* (*Adjusted Median*). Specifically, let $z_i = 1$ if *Weighted Mean* (*Adjusted Median*) is successful at predicting the earnings surprise, and $z_i = 0$ otherwise. The success probability of a single trial is denoted by $p = prob(z_i = 1)$. Denoting z as the number of successes in n trials, i.e., $z = \sum_{i=1}^{n} z_i$, it follows that z is distributed as a $\text{Binomial}(n, p)$.

To test whether the success probability is greater than 50%, two tests can be used. First, for the null hypothesis of $p = p_0 = 0.5$, the following test statistic can be used,

$$\tau = \frac{\hat{p} - p_0}{\sqrt{\hat{p}(1-\hat{p})/n}}, \text{ where } \hat{p} = \frac{z}{n}, \text{ and } \tau \sim N(0,1).$$

Second, the complementary cumulative distribution function (CCDF), defined as $p(z > z_0) = 1 - F(z_0)$, tells how often the random variable z is above a particular level z_0. For given z_0 and n, if CCDF is small for $p_0 = 0.5$, it means observing z_0 successes in n trials with success probability 1/2 is a low probability event. If $z = z_0$ actually occurs, it is unlikely for the null hypothesis $p_0 = 0.5$ to be true.

Using the aforementioned method, the ability of *Weighted Mean* (*Adjusted Median*) to forecast earnings surprises is analyzed using five out-of-sample quarters. The earnings surprises forecasting results of *Weighted Mean* (*Adjusted Median*) for 2004:Q1 are shown in Table 3.4 . The results for 2004:Q2 to 2005:Q1 are qualitatively similar to 2004:Q1, therefore, they are not presented here to avoid redundancy. For each of the five quarters studied, *Weighted Mean* (*Adjusted Median*) are constructed using estimation results from sub-samples ending at the previous quarter. Within each quarter, the results are further divided into sub-groups by the number of analysts that follows a firm. Table 3.4 lists the number of firms in each sub-group, the mean of z_i, the p-value under the null hypothesis $p_0 = 0.5$ (indicated by mean's superscript *), and CCDF with $p_0 = 0.5$. For example, for Q1, 2004, there are 305 firms that have exactly 3 analysts following them. The forecasting success rate for *Weighted Mean* is 56.4%, and the null hypothesis of $p_0 = 0.5$ is rejected at the 5% significance level. The probability of observing more than 169 ($305 \times 56.4\%$) successes out of 305 trials is 0.015. Therefore, the null hypothesis $p_0 = 0.5$ is also rejected at the 5% significance level by the CCDF test. The forecasting success rate for *Adjusted Median* is 58.0% , and the null hypothesis of $p_0 = 0.5$ is rejected at the 1% significance level. The probability of observing more than 305*58.0% successes out of 305 trials is 0.002. Therefore, the null hypothesis $p_0 = 0.5$ is also rejected at the 1% significance level by the CCDF test.

Table 3.4 Earnings Surprises Forecasting Results

		Weighted Mean		Adjusted Median	
No. of Analysts	No. of Companies	Mean	CCDF	Mean	CCDF
1	813	0.566***	0.000	0.566***	0.000
2	431	0.568***	0.002	0.575***	0.001
3	305	0.564**	0.015	0.580***	0.002
4	190	0.553	0.064	0.495	0.586
5	168	0.589**	0.012	0.577**	0.018
6	103	0.602**	0.024	0.602**	0.024
7	93	0.570	0.073	0.559	0.107
8	82	0.488	0.544	0.488	0.544
9	66	0.591	0.054	0.576	0.088
10	43	0.605	0.063	0.605	0.063
>10	168	0.613***	0.002	0.583**	0.018

Note: *** significant at 1% level. ** significant at 5% level. * significant at 10% level.

In the Bayesian framework, not only can one get the point estimate of F_{aft}^{t+1*}, one can also get the probability distribution of F_{aft}^{t+1*}. The sampling distribution of an out-of-sample F_{aft}^{t+1*} would be an acceptable predictive distribution if Γ were known. However, without knowledge of Γ, this can not be used. In its place is the Bayesian predictive probability distribution,

$$p(F_{aft}^{t+1*} \mid y) = \int_{\Gamma} p(F_{aft}^{t+1*} \mid y, \Gamma) p(\Gamma \mid y) d\Gamma \, .$$

Similarly, the sampling distribution of *Weighted Mean* could also be derived. Figure 3.3 plots the Bayesian predictive density of *Weighted Mean* for Yahoo's 2005:Q1. Also plotted in Figure 3.3 are actual EPS (henceforth actual) and consensus *Mean* EPS (henceforth consensus). The consensus is just the simple mean of relevant forecasts in the sample.

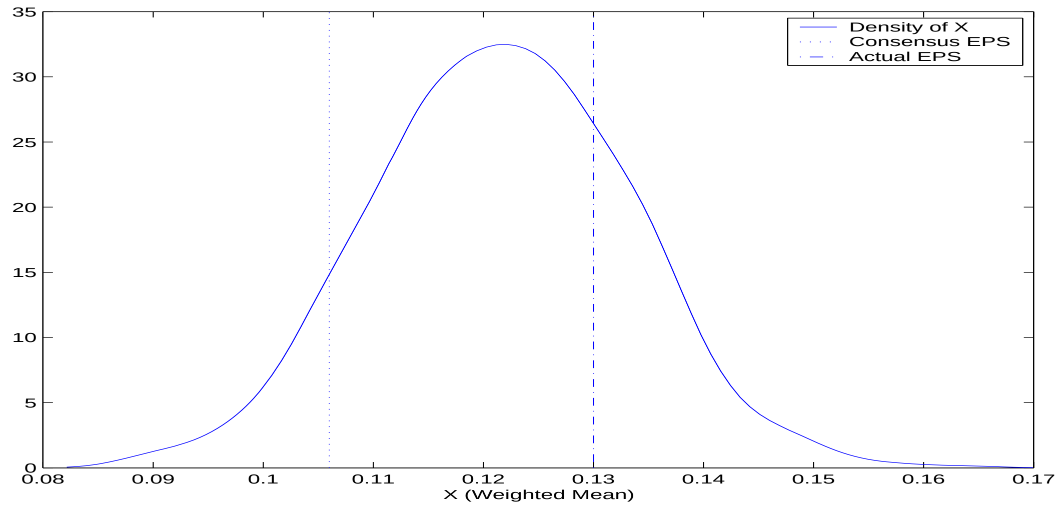


Figure 3.3 The Bayesian Predictive Density

Since the density of weighted mean gives the probabilities associated with different values of *Weighted Mean*, the relative position of the consensus over the density of *Weighted Mean* offers hints about whether the consensus will under-estimate or over-estimate the actual. If the majority of *Weighted Mean* 's mass is to the right of the consensus (as shown in Figure 3.3), it means that the better analysts identified by the proposed method think the actual will be larger than consensus; therefore, the consensus will likely under-estimate actual. For the case of Figure 3.3, 91.1% of the draws of *Weighted Mean* are larger than the consensus. Since the actual EPS is higher than the consensus, in this instance, the density of *Weighted Mean* yields the correct prediction of earnings surprise.

The earnings surprises forecasting results of all available companies for 2005:Q1 are shown in Table 3.5. Table 3.5 is further broken into whether the densities of *Weighted Mean* predict positive surprises or negative surprises. The prediction confidence refers to the percentages of *Weighted Mean*'s mass to the right or left of the consensus. For example, 95%-99% prediction confidence means that 95%-99% of the draws of *Weighted Mean* are larger than the consensus for positive surprises in Panel A, or are smaller than the consensus for negative surprises in Panel B. For the prediction of positive earnings surprises, at the 95%-99% prediction confidence level, there are 49 successes out of 66 predictions. The success rate p is 0.7424 with a standard deviation of 0.0538. For the prediction of negative earnings surprises, at the 95%-99% prediction confidence level, there are 37 successes out of 56 predictions. The success rate p is 0.6607 with a standard deviation 0.0633. Both success rates are statistically higher than 0.5 and are quantitatively large enough to be economically significant.

It is interesting to note that for the confidence levels 50%-75%, the success rates are close to 0.5. However, for the confidence levels greater than 75%, the success rates for positive and negative earnings surprises are 0.7009 and 0.6480 respectively. As the confidence levels get higher, the success rates generally get larger. In the extreme, the success rate for predicting positive earnings surprises with greater than 99.5% confidence level is 0.8611.

Table 3.5 Predicting Earnings Surprises Using Bayesian Predictive Densities

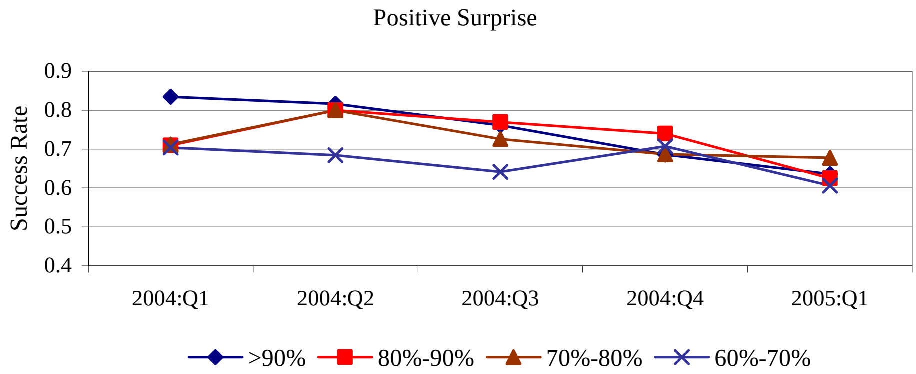
Panel A: Predicting consensus will under-estimate actual (positive earnings surprise)

Prediction Confidence	Number of Predictions	Number of Successes	Success Rate (P)	Std. Dev. of P
>99.5%	36	31	0.8611	0.0576
99%-99.5%	14	12	0.8571	0.0935
95%-99%	66	49	0.7424	0.0538
90%-95%	84	53	0.6310	0.0527
85%-90%	68	45	0.6618	0.0574
80%-85%	102	72	0.7059	0.0451
75%-80%	98	66	0.6735	0.0474
>75%	468	328	0.7009	0.0212
50%-75%	688	377	0.5480	0.0190

Panel B: Predicting consensus will over-estimate actual (negative earnings surprise)

Prediction Confidence	Number of Predictions	Number of Successes	Success Rate (P)	Std. Dev. of P
>99.5%	35	25	0.7143	0.0764
99%-99.5%	8	7	0.8750	0.1169
95%-99%	56	37	0.6607	0.0633
90%-95%	66	47	0.7121	0.0557
85%-90%	52	30	0.5769	0.0685
80%-85%	75	46	0.6133	0.0562
75%-80%	83	51	0.6145	0.0534
>75%	375	243	0.6480	0.0247
50%-75%	647	340	0.5255	0.0196

The present study also estimated the success rate for earnings surprises using the Bayesian predictive density for 2004:Q1 to 2004:Q4. The results are presented in Figure 3.4 and Figure 3.5. In Figure 3.4, the predictions are based on the estimated results using the sample ending at 2003:Q4. For positive surprises, the success rates are generally higher than 0.6 and decrease as the forecast horizon gets longer. For negative surprises, the success rates are generally higher than 0.5. Figure 3.4 also shows that the higher the confidence level, the higher the success rates.



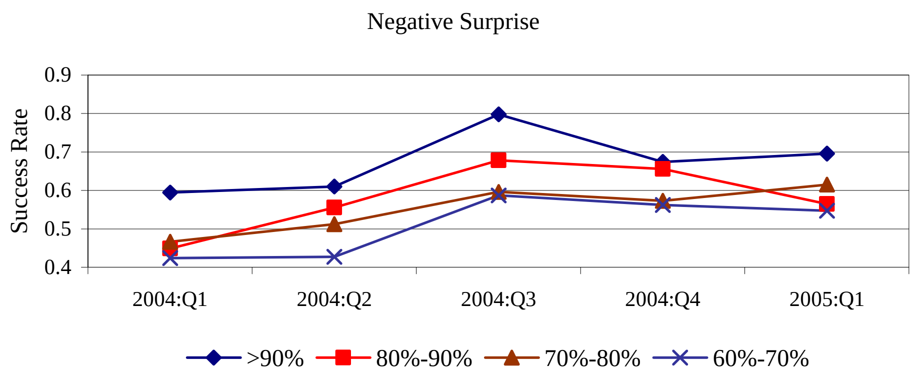
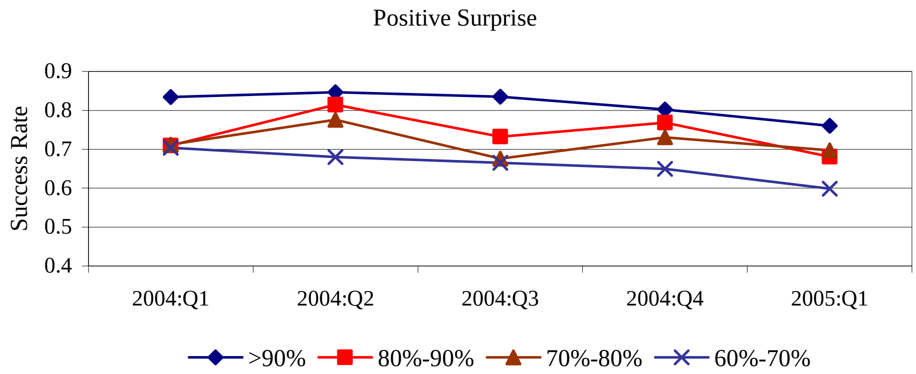


Figure 3.4 Earnings Surprises Predictions Based on 2003:Q4

In Figure 3.5, the predictions are based on the estimated results using the sample ending at the previous quarter. When comparing Figure 3.4 with Figure 3.5, the shapes of the graphs do not vary significantly. This indicates that even some forecast horizons do not utilize all available information, the decrease in forecasting performance is not very significant. This shows that analysts' abilities are fairly stable over time. For positive surprises, the success rates are generally higher than 0.6. There is less variation in success rates when compared with those in Figure 3.4. For negative surprises, the success rate are generally higher than 0.5. As in Figure 3.4, Figure 3.5 also shows that the higher the confidence level, the higher the success rate.



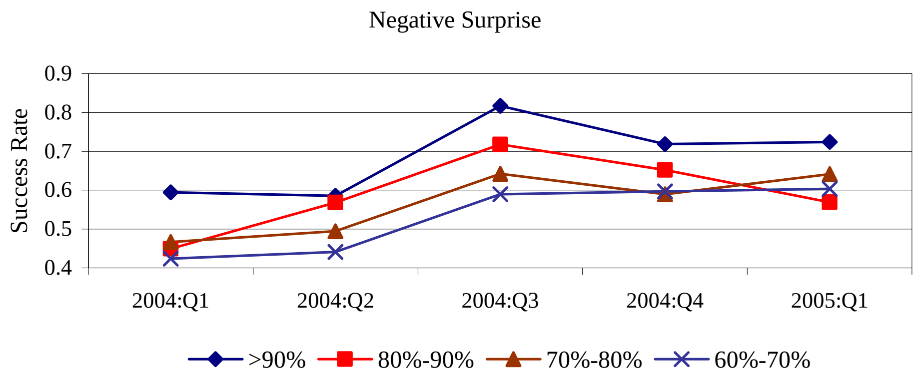


Figure 3.5 Earnings Surprises Predictions Based on Previous Quarter

In summary, this subsection shows that the *Weighted Mean* derived here could help to predict whether the original forecast will under- or over-estimate the actual EPS. Moreover, the density of *Weighted Mean* could also help to quantify the level of certainty about the predictions. The levels of certainty are especially useful for practical applications. Since it has been documented in the finance and accounting literatures that the stock market reacts to earnings announcements,[9] when constructing portfolios based on the earnings surprises

---

[9] For example, Sultan (1994) finds that unexpected earnings can be used as a discriminator between stocks that performed relatively well and stocks that performed relative poorly in Japan. Brown and Jeong (1998) show that an earnings surprise predictor is effective in selecting stocks from S&P 500 firms. Dische and Zimmermann (1999) report that abnormal returns can be earned from the portfolio of Swiss stocks exhibiting

predictions, one could just focus on the predictions associated with a high level of certainty, e.g. greater than 75%.

3.8 Conclusion

Using a three-level Bayesian hierarchical model, consistent with previous studies, the present study shows that as a population, analysts' forecasts are biased and inefficient. Analysts are systematically optimistic and their forecasts are too extreme. As for efficiency, the results show that if an analyst over-estimates EPS in one period, he will tend to over-estimate EPS in the next period, and vice versa. There is also evidence that if a firm's EPS in the current quarter is greater than last quarter's, analysts tend to under-estimate next quarter's EPS, and vice versa. The Bayesian hierarchical model allows us to avoid making scaling transformations to the original data, and to relax the unrealistic assumption of analyst homogeneity. The main contribution of the present study is that, as a result of the proposed method, we are able to analyze forecasts at the individual analyst level. This allows us to identify that there is considerable heterogeneity in the degrees of analysts' bias and inefficiency. The forecasts of some analysts, especially those of some firms that these analysts follow, can be regarded as unbiased and efficient.

The present study also adjusts for biases in analysts' original forecasts. From the adjusted forecasts, *weighted mean* and *adjusted median* are derived. The proposed *weighted mean* and *adjusted median* measures are shown to be able to forecast earnings surprises with success rates that are statistically higher than 0.5. The levels of certainty about the

---

the most positive earnings revision. Conroy, Eades, and Harris (2000) find that stock prices are significantly affected by earnings surprises in Japan. Mozes (2000) shows that the strategy of buying stocks on the basis of positive forecasted earnings surprises is more profitable for value firms than for growth firms.

predictions could also be quantified. As the confidence levels get higher, the success rates generally get larger. When the prediction confidence levels are greater than 75%, the success rates are generally higher than 0.6.

3.8 References

Amir, Eli and Yoav Ganzach (1998), Overreaction and underreaction in analysts' forecasts. Journal of Economic Behavior and Organization, 37, 3, 333-347.

Ali, A., A. Klein, and J. Rosenfeld (1992), Analysts' use of information about permanent and transitory earnings components in forecasting annual EPS, Accounting Review, 67, 1, 183-198.

Abarbanell, J.S. (1991), Do analysts' earnings forecasts incorporate information in prior stock price changes? Journal of Accounting and Economics. 14, 2, 147-165.

Abarbanell, J. and V. Bernard (1992), Tests of analysts' overreaction/underreaction to earnings information as an explanation for anomalous stock price behavior. Journal of Finance, 47, 1181-1207.

Brown, Lawrence D., Robert L. Hagerman, Paul A. Griffin, and Mark E. Zmijewski (1987), Security analyst superiority relative to univariate time-series models in forecasting quarterly earnings. Journal of Accounting and Economics 9, 61-87.

Brown, L. D. and S. W. Jeong (1998), Profiting from predicting earnings surprise. Journal of Financial Statement Analysis, 57-66.

Clement, M. (1999), Analyst forecast accuracy: do ability, resources, and portfolio complexity matter? Journal of Accounting and Economics, 27, 285-303.

Constantinou, Ph. Constantina, William P. Forbes, and Len Skerratt (2003), Analyst underreaction in the United Kingdom. Financial Management, Summer, 93-106.

Conroy, R.M., K.M. Eades, and R.S. Harris (2000), A test of the relative pricing effects of dividends and earnings: evidence from simultaneous announcements in Japan. Journal of Finance, 1199-1227.

Daniel A. Cohen and Thomas Z. Lys (2003), A note on analysts' earnings forecast errors distribution. Journal of Accounting and Economics, 36, 147-164

De Bondt, Werner F. M., and Richard H. Thaler (1990), Do security analysts overreact? American Economic Review 80, 52-57.

Dische, A. and H. Zimmermann (1999), Consensus forecasts of corporate earnings changes and the performance of Swiss stocks. Journal of Investing Spring 19-26.

Easterwood, J. and S. Nutt (1999), Inefficiency in analysts' earnings forecasts: systematic misreaction or systematic optimism? Journal of Finance, 54, 5, 1777-1797.

Elliot, J., D. Philbrick, and C. Wiedman (1995), Evidence from archival data on the relation between security analysts' forecast errors and prior forecast revisions, Contemporary Accounting Research, 11, 919-938.

Goldstein, Harvey (2003), Multilevel statistical models. Kendall's library of statistics 3, Arnold, London.

Hong, Harrison, Jeffrey D. Kubik (2003), Analyzing the analysts: career concerns and biased earnings forecasts. Journal of Finance 58, 1, 313-351.

Jacob, J., T. Lys, M. Neale (1999), Expertise in forecasting performance of security analysts. Journal of Accounting and Economics, 28, 51-82.

Kothari, S.P. (2001), Capital markets research in accounting. Journal of Accounting and Economics 31, 105-232.

Kothari, S.P., J.S. Sabino and T. Zach (2005), Implications of survival and data trimming for tests of market efficiency, Journal of Accounting and Economics, Vol. 39, 1, 129-161.

Koop, Gary, Dale J. Poirier, and Justin L. Tobias (2006), Bayesian econometrics. Cambridge University Press. Forthcoming.

Loh, Roger K., and G. Mujtaba Mian (2006), Do accurate earnings forecasts facilitate superior investment recommendations? Journal of Financial Economics. Forthcoming.

Lys, T. and S. Sohn (1990), The association between revisions of financial analysts' earnings forecasts and security-price changes, Journal of Accounting and Economics 13, 4, 341-363.

Mikhail, M., Beverly R. Walther; Richard H. Willis (1997), Do security analysts improve their performance with experience? Journal of Accounting Research, 35, 131-166.

Mozes, H.A. (2000), The role of value in strategies based on anticipated earnings surprise. Journal of Portfolio Management Winter, 54-62.

Sinha, P., Lawrence D. Brown, and S. Das (1997), A re-examination of financial analysts' differential earnings forecast accuracy. Contemporary Accounting Research 14, 1-42.

Sultan, A.T. (1994), Earnings surprise in Japan. Journal of Investing Winter, 32-38.

Womack, Kent L. (1996), Do brokerage analysts recommendation have value? Journal of Finance 51, 137-167.

3.9 Appendix

This appendix describes the posterior distributions of the parameters in regression (3.5).

Complete posterior conditional for θ_{af} :

$$p(\theta_{af} \mid \Gamma_{-\theta_{af}}, y) \sim N(D_{\theta_{af}} d_{\theta_{af}}, D_{\theta_{af}}),\ \text{where}$$

$$D_{\theta_{af}} = (X'_{af} X_{af} / \sigma^2_{af} + \Omega_a^{-1})^{-1},$$

$$d_{\theta_{af}} = X'_{af} y_{af} / \sigma^2_{af} + \Omega_a^{-1} \theta_a .$$

Complete posterior conditional for σ_{af}^2 :

$$p(\sigma_{af}^2 \mid \Gamma_{-\sigma_{af}^2}, y) \sim IG\left(\frac{N_{af}}{2}+a, \left[\frac{1}{2}(y_{af}-X_{af}\theta_{af})^{'}(y_{af}-X_{af}\theta_{af})+\frac{1}{b}\right]^{-1}\right).$$

Complete posterior conditional for θ_a:

$$p(\theta_a \mid \Gamma_{-\theta_a}, y) \propto \left\{\prod_{f=1}^{F_a} p(\theta_{af} \mid \theta_a, \Omega_a)\right\} p(\theta_a \mid \theta_0, \Sigma)$$

Since the second stage of the model specifies $p(\theta_{af} \mid \theta_a, \Omega_a)$ as *iid* , the following equation applies,

$$\begin{bmatrix} \theta_{a,1} \\ \theta_{a,2} \\ \vdots \\ \theta_{a,F_a} \end{bmatrix} = \begin{bmatrix} I_2 \\ I_2 \\ \vdots \\ I_2 \end{bmatrix} \theta_a + \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{F_a} \end{bmatrix},$$

or equivalently,

$$\tilde{\theta}_{af} = \tilde{I}\theta_a + \tilde{\mu}, \text{ where,}$$

$$\tilde{\theta}_{af} = \begin{bmatrix} \theta_1^{'} & \theta_2^{'} & \cdots & \theta_{F_a}^{'} \end{bmatrix}^{'},$$

$$\tilde{I} = \begin{bmatrix} I_2 & I_2 & \cdots & I_2 \end{bmatrix}^{'},$$

$$\tilde{\mu} = \begin{bmatrix} \mu_1^{'} & \mu_2^{'} & \cdots & \mu_{F_a}^{'} \end{bmatrix}^{'},$$

$$\text{and } E(\tilde{\mu} \bullet \tilde{\mu}^{'}) = I_{F_a} \otimes \Omega_a .$$

Expressed in the above form, the posterior of θ_a can be shown to be normal and has the following form.

$$p(\theta_a \mid \Gamma_{-\theta_a}, y) \sim N(D_{\theta_a} d_{\theta_a}, D_{\theta_a}), \text{ where}$$

$$D_{\theta_a} = (\tilde{I}^{'}(I_{F_a} \otimes \Omega_a^{-1})\tilde{I} + \Sigma^{-1})^{-1} = (F_a \Omega_a^{-1} + \Sigma^{-1})^{-1}$$

$$d_{\theta_a} = (\tilde{I}^{'}(I_{F_a} \otimes \Omega_a^{-1})\overline{\theta}_{af} + \Sigma^{-1}\theta_0) = (F_a \Omega_a^{-1}\overline{\theta}_{af} + \Sigma^{-1}\theta_0), \text{ where } \overline{\theta}_{af} = \frac{\sum_{f=1}^{F_a}\theta_{af}}{F_a}.$$

Complete posterior conditional for Ω_a^{-1}:

$$p(\Omega_a^{-1} \mid \Gamma_{-\Omega_a^{-1}}, y) \sim Wishart(\left[\sum_{f=1}^{F_a} (\theta_{af} - \theta_a)(\theta_{af} - \theta_a)' + \rho_\Omega R_\Omega\right]^{-1}, F_a + \rho_\Omega)$$

Complete posterior conditional for θ_0:

$$p(\theta_0 \mid \Gamma_{-\theta_0}, y) \sim N(D_{\theta_0} d_{\theta_0}, D_{\theta_0}), \text{ where}$$

$$D_{\theta_0} = (\tilde{I}'(I_A \otimes \Sigma^{-1})\tilde{I} + C^{-1})^{-1} = (A\Sigma^{-1} + C^{-1})^{-1}$$

$$d_{\theta_0} = (\tilde{I}'(I_A \otimes \Sigma^{-1})\overline{\theta}_a + C^{-1}\eta) = (A\Sigma^{-1}\overline{\theta}_a + C^{-1}\eta), \text{ where } \overline{\theta}_a = \frac{\sum_{a=1}^{A} \theta_a}{A}.$$

Complete posterior conditional for Σ^{-1}:

$$p(\Sigma^{-1} \mid \Gamma_{-\Sigma^{-1}}, y) \sim Wishart(\left[\sum_{a=1}^{A} (\theta_a - \theta_0)(\theta_a - \theta_0)' + \rho_\Sigma R_\Sigma\right]^{-1}, A + \rho_\Sigma)$$

CHAPTER 4. EVALUATING THE INFORMATION CONTENT OF FORECASTS

4.1 Abstract

The present study proposes a new approach to compare forecasts' information contents. Following recent developments in the forecasting literature, it regards forecasts as predictors, and derives next period's expected value of the variable being forecasted conditional on alternative information sets using the Kalman filter. Forecasts that contain more information will lead to a smaller variance of deviations between actual values and expected values. The relative magnitude of the above variance is regarded as the measure of the relative information contents of competing forecasts. The present study also proposes a way to measure the information content of forecasts from the same source without competing forecasts, which could not be determined by previous methods. The advocated measures are computed for a well-known data set and yield different conclusions from those drawn by the previous literature's regression-based measures. The proposed measures do not suffer from the multicollinearity problem that could affect previous regression-based measures. Furthermore, they are derived from well-defined information sets. The flexibility researchers enjoy in constructing the information sets allows the proposed measures to be applied to various situations.

4.2 Introduction

Forecasts of future values of macroeconomic variables, company earnings, etc., are widely perceived to provide useful information about economic agents' expectations. One question of particular interest is how to determine forecasts' information contents of actual values. Forecasts' information content and accuracy are inextricably linked. A forecast that contains more information about the actual value should have a smaller forecast error than a less informative forecast. However, various forecast accuracy measures based on root mean squared errors (RMSE) may not be direct measures of forecasts' information content. For example, if the RMSEs of two forecasts are so close that the differences are not economically meaningful, little can be said about which one contains more information. In addition, the RMSE framework assumes forecasters have symmetric quadratic loss functions. In reality, forecasters may have other types of loss functions. There are cases when the forecast about the direction of change is at least as important as the forecast of the actual level, such as variables related to futures and options. Forecasters could potentially maximize trading profits by correctly predicting the direction of change, regardless of the magnitude of mean squared errors.

In contrast to the forecast accuracy literature, which contains literally hundreds of forecast accuracy comparisons, there are few papers that explicitly compare the information contents of competing forecasts. Notable examples of the latter include Fair and Shiller (1989, 1990), and Romer and Romer (2000). Fair and Shiller (1989, 1990) examined whether one model's forecast of real GNP carries different information from another model's forecast by regressing the actual change in real GNP on the forecasted changes from the two models. Let y_{t+s} denote the actual value of variable y at time $t+s$. Let x_{1t}^{t+s} (from model 1) and x_{2t}^{t+s} (from model 2) denote two competing forecasts of y_{t+s} as of time t. Fair and Shiller (1989, 1990) run the following regression,

$$y_{t+s} - y_t = \alpha + \beta(x_{1t}^{t+s} - y_t) + \gamma(x_{2t}^{t+s} - y_t) + \varepsilon_{t+s}\,. \tag{4.1}$$

Fair and Shiller (1990) regarded the hypothesis $H_0 : \beta = 0$ as the hypothesis that forecast x_{1t}^{t+s} contains no information, in addition to the constant term and forecast x_{2t}^{t+s}, relevant to forecasting the s-period-ahead actual value y_{t+s}; and the hypothesis $H_0 : \gamma = 0$ as the hypothesis that forecast x_{2t}^{t+s} contains no information, in addition to the constant term and forecast x_{1t}^{t+s}, relevant to forecasting the s-period-ahead actual value y_{t+s}. Fair and Shiller (1990) used actual GNP changes and forecasted changes, instead of GNP level and forecasted levels in equation (1) because they suspected that GNP may be an integrated process. When studying whether the Federal Reserve has additional information compared to commercial forecasters, Romer and Romer (2000) used levels in equation (4.1), as they believed that the inflation rate is a stationary process. Their interpretation of β and γ is similar to Fair and Shiller (1990).

Another related strand of literature consists of forecast encompassing studies. Forecast encompassing implicitly compares the information contents of competing forecasts. Forecast encompassing tests evaluate whether competing forecasts can be combined into a better forecast. Such tests are usually implemented by regressing the actual level of y_{t+s} (or the actual change) on the forecasts (or the forecasted changes) from two models (e.g., Chong and Hendry (1986), Ericsson (1993), Stock and Watson (1999), Ang, Bekaert and Wei (2005)),

$$y_{t+s} = \beta x_{1t}^{t+s} + \gamma x_{2t}^{t+s} + \varepsilon_{t+s}, \text{ subject to } \beta + \gamma = 1 \tag{4.2}$$

Equation (4.2) is a restricted version of equation (4.1) with $\alpha = 0$ and $\beta + \gamma = 1$. Conditional on $\beta + \gamma = 1$, if $\beta \neq 0$ and $\gamma = 0$, the first model encompasses the second model. According to Chong and Hendry (1986) and Clements and Hendry (1993), one forecast encompasses another forecast if the weight assigned to the first forecast is not significantly different from one and the weight assigned to the second forecast is not significantly different from zero when combining the two forecasts. If one forecast is found to encompass another forecast, then the encompassed forecast does not contain extra information in addition to the encompassing forecast, at least in the sense of linear combination.

One disadvantage of equations (4.1) and (4.2) is that x_{1t}^{t+s} is often highly correlated with x_{2t}^{t+s} because of the following two reasons. First, they both contain public information. Second, in many cases, one model's forecaster knows the other model's forecast if the two forecasts are not announced simultaneously. Hence, two models' forecasts could be further correlated due to overlapping private information. Therefore, equations (4.1) and (4.2) often likely suffer from multicollinearity problems. Collinear variables do not provide enough information to estimate their separate effects. Some of the variances, standard errors, and covariances of the OLS estimators may be large as seen in Romer and Romer (2000).

The present study proposes a new approach to measure the information content of competing forecasts. The advocated approach is based on the results of Pastor and Stambaugh (2006)'s *predictive system*, a state-space model that they first applied to the problem of stock returns predictability. They focused on how to derive better estimates of expected stock returns. Pastor and Stambaugh (2006) found that the predictive system could deliver different and substantially more precise estimates than the standard regression approach when used to predict stock returns.

Pastor and Stambaugh (2006) decomposed y_{t+1} as,

$$y_{t+1} = \mu_t + u_{t+1}, \tag{4.3}$$

where $\mu_t \equiv E(y_{t+1} | D_t)$ is the expectation of y_{t+1} conditional on information set D_t at time t, u_{t+1} is the un-forecasted shock to y from t to $t+1$ and has mean zero conditional on information set D_t. Regarding μ_t as the unobserved state, the value of μ_t can be derived through the Kalman filter (Kalman (1960)). The estimated μ_t is considered as the best estimate of y_{t+1} by the researcher (the general public) based on information set D_t. The information set D_t consists of the collection of variables that the predictive system uses to derive μ_t. Because of the flexibility of the predictive system, researchers could arbitrarily change the variables in D_t. Each estimated μ_t will be unique to the D_t used to derive it. Depending on D_t, the accuracy of estimated μ_t will be different. If D_t contains variables that are highly correlated with y_{t+1}, the estimated μ_t will be highly correlated with y_{t+1}. Consequently, the un-forecasted part of y_{t+1}, u_{t+1}, will be small. Therefore, the more information $\{D_t\}_{t=1}^{T}$ contains about $\{y_{t+1}\}_{t=1}^{T}$, the smaller will be the variance of $\{u_{t+1}\}_{t=1}^{T}$.

By focusing on the variance of $\{u_{t+1}\}_{t=1}^{T}$, researchers could use the predictive system to study the information content of forecasts. Specifically, the variances of $\{u_{t+1}\}_{t=1}^{T}$ associated with different information sets can be calculated. Since the information sets' composition are clearly defined in terms of the variables included in it, the researcher can choose various information sets by including extra variables or excluding existing variables. The ratio of the variances of $\{u_{t+1}\}_{t=1}^{T}$ can be regarded as the measure of the relative information content of different information sets. The differences in the information contents can be attributed to the differences in the variable composition of information sets which are controlled by the researcher.

The contribution of the present study is mainly methodological. It proposes new measures that can determine the relative information contents of forecasts explicitly accounting for the conditioning information. The proposed measures are derived from clear and well-defined information sets. The flexibility researchers enjoy in constructing the information sets allows the proposed measures to be applied to various situations. By contrast, previous studies' regression-based measures do not have clear and well-defined information sets. They could only compare the information contents of competing forecasts and could not determine the information content of a single set of forecasts. The measures proposed in the present study do not suffer from multicollinearity problems that could affect the regression-based measures. The proposed measures are applied to the data used by Romer and Romer (2000). The empirical application shows that both measures perform reasonably well. It is found that the Federal Reserve is better than commercial forecasters in forecasting inflation rates. But its informational advantage in only confined to the very short term. The 2- to 4-quarter-ahead forecasts of both the Federal Reserve and commercial forecasters do not offer much extra information about future inflation rates in addition to past inflation rates. Whereas Romer and Romer (2000) believed that the Federal Reverse forecasts are good for all horizons. The present study also extends the forecasting literature by applying Pastor and Stambaugh (2006)'s predictive system to the literature on forecasts' information content.

The remainder of the paper is organized as follows. Section 3 outlines the predictive system and the information content measures. Section 4 gives an application of the information content measures. Section 5 concludes.

4.3 Measures of Information Contents

4.3.1 The predictive system and its application to survey forecasts

Pastor and Stambaugh (2006) originally applied the predictive system to the problem of stock returns predictability. However, as shown here, the predictive system is also well-suited for the analysis of survey forecasts. The goal of the predictive system is to estimate the states of a dynamic system ($\{\mu_t\}_{t=1}^T$) from a series of noisy measurements ($\{y_t\}_{t=1}^T$, and $\{x_t^{t+1}\}_{t=1}^T$ if available) by the Kalman filter. The Kalman filter is very powerful in several aspects: it supports estimations of past, present, and future states, and it can do so even when the precise nature of the modeled system is unknown. Regarding forecasts as noisy measurements of y_{t+1}, the predictive system can efficiently estimate the states of $\{y_{t+1}\}_{t=1}^T$, $\{\mu_t\}_{t=1}^T$. The quality of $\{\mu_t\}_{t=1}^T$ and the other parameters in the predictive system can shed light on the quality of forecasts. The main features of the predictive system are discussed in the following paragraphs.

Assuming μ_t obeys the first-order autoregressive process,

$$\mu_{t+1} = \alpha + B\mu_t + w_{t+1}, \tag{4.4}$$

the Kalman filter technique can be used to derive the unobserved states $\{\mu_t\}_{t=1}^T$ with equations (4.3) and (4.4) alone. In the language of the Kalman filter, equation (4.3) is regarded as the observation equation and equation (4.4) is regarded as the state equation. Since forecasters base their forecasts on the variables related to $\{\mu_t\}_{t=1}^{T}$, the innovations of forecasts contain useful information about the change of these variables which could help to infer the value of $\{\mu_t\}_{t=1}^{T}$, even if forecasts are biased. Therefore, forecasts, as well as variables that are correlated with $\{\mu_t\}_{t=1}^{T}$, should be included in the Kalman filter. Forecasts are assumed to follow a first-order autoregressive process,

$$x_{t+1}^{t+2} = \theta + Ax_t^{t+1} + v_{t+1}\,. \tag{4.5}$$

Equations (4.3), (4.4), and (4.5) can be combined to formed a predictive system for μ_t:

$$y_{t+1} = \mu_t + u_{t+1} \tag{4.6}$$

$$x_{t+1}^{t+2} = \theta + Ax_t^{t+1} + v_{t+1} \tag{4.7}$$

$$\mu_{t+1} = \alpha + B\mu_t + w_{t+1} \tag{4.8}$$

The residuals in the system are assumed to be distributed identically and independently across t as,

$$\begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix} \sim N(0,\Sigma) = N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv} & \Sigma_{uw} \\ \Sigma_{vu} & \Sigma_{vv} & \Sigma_{vw} \\ \Sigma_{wu} & \Sigma_{wv} & \Sigma_{ww} \end{bmatrix}\right) \tag{4.9}$$

The predictive system is a version of a state-space model in which there is non-zero correlation among the disturbances. The value of μ_t is unobservable, but the predictive system implies a value of $E(\mu_t \mid D_t) = E(y_{t+1} \mid D_t)$, where $D_t = \{y_t, y_{t-1}, \cdots, x_t^{t+1}, x_{t-1}^{t+2}, \cdots\}$, the history of actual values and forecasts observed through time t.

The composition of information set D_t depends on the specification of equation (4.7). Researchers could change equation (4.7) to include or exclude specific variables. Using the Kalman filter, $E(y_{t+1} \mid D_t)$ can be written as the unconditional mean of y plus a linear combination of past un-forecasted shocks $\{u_t, u_{t-1}, u_{t-2}, \cdots\}$, and innovations in the forecasts $\{v_t, v_{t-1}, v_{t-2}, \cdots\}$. Specifically, the expected value of y_{t+1} conditional on $D_t = \{y_t, y_{t-1}, \cdots, x_t^{t+1}, x_{t-1}^{t+2}, \cdots\}$ is given by,

$$E(y_{t+1} \mid D_t) = E(y) + \sum_{s=0}^{\infty} (\psi_s u_{t-s} + \phi_s v_{t-s}),$$

where ψ_s and ϕ_s are functions of the parameters in (4.6) through (4.9). In essence, $E(\mu_t \mid D_t)$ will be a linear function of variables in the set D_t, i.e., $E(y_{t+1} \mid D_t) = f(y_t, y_{t-1,} \cdots, x_t^{t+1}, x_{t-1}^{t+2}, \cdots)$. The forecast innovation term v_t, which is the change from $\theta + Ax_{t-1}^t$ to x_t^{t+1}, summarizes the changes happened during time period t that the forecaster thinks are relevant for the prediction of the actual value y_{t+1}.

The parameters in the predictive system are estimated using a Bayesian approach. The Bayesian approach has several advantages over frequentist alternatives such as the maximum likelihood method. First, the former incorporates parameter uncertainty as well as uncertainty about the path of the unobservable state $\{\mu_t\}_{t=1}^T$. Second, it allows posterior distributions to be easily calculated for arbitrary functions of parameters. Non-informative priors are employed for both (θ, A, α, B) and Σ. The posterior distributions are derived using Gibbs sampling. In each step of the MCMC chain, the parameters (θ, A, α, B) and Σ are first drawn conditional on the current draw of $\{\mu_t\}_{t=1}^T$. Then the forward filtering, backward sampling algorithm developed by Carter and Kohn (1994) and Fruhwirth-Schnatter (1994) is used to draw the time series of $\{\mu_t\}_{t=1}^T$ conditional on (θ, A, α, B) and Σ. The details of each step are given in the Appendix.

4.3.2 Measuring the information content of one set of forecasts

The predictive system regards μ_t as the forecast of y_{t+1} conditional on information set D_t for $t = 1, \cdots, T$. By focusing on the variance of the un-forecasted shocks $\{u_{t+1}\}_{t=1}^T$, researchers can use the predictive system to study the information content of forecasts. Since the composition of the predictive system's information set is controllable by the researcher, various information sets can be constructed by including extra variables or excluding existing variables. The variances of $\{u_{t+1}\}_{t=1}^T$ associated with different information sets can be calculated. The ratio of the variances of $\{u_{t+1}\}_{t=1}^{T}$ can be regarded as the measure of the relative information content of different information sets.

Consider for example, the case when researchers only have forecasts from the same source without competing forecasts. In this instance, equation (4.7) could be lagged by one period,

$$x_t^{t+1} = \theta + Ax_{t-1}^{t} + v_{t+1}. \tag{4.10}$$

In the predictive system consisting of equations (4.6), (4.8), and (4.10), the information set is $D_t = \{y_t, y_{t-1}, \cdots, x_{t-1}^{t}, x_{t-2}^{t-1}, \cdots\}$, which contains only past histories of actual earnings and forecasts, but not the current forecast x_t^{t+1} for y_{t+1}. In essence, x_t^{t+1} is considered unknown as of time t, x_{t-1}^{t} is considered unknown as of time $t-1$, etc. In contrast, in the predictive system consisting of equations (4.6), (4.8), and (4.7), the information set is $D_t = \{y_t, y_{t-1}, \cdots, x_t^{t+1}, x_{t-1}^{t}, \cdots\}$, which contains the current forecast x_t^{t+1} too. Therefore, the information set D_t used to calculate $E(y_{t+1} \mid D_t)$ in the predictive system consisting of equations (4.6), (4.8), and (4.10) is a subset of the information set D_t used in the predictive system consisting of equations (4.6), (4.8), and (4.7). The only difference between the two information sets is that x_t^{t+1} is included in the latter but not in the former. Hence, any differences between two predictive systems' estimates could only come from the additional information contained in the current forecast $\{x_t^{t+1}\}_{t=1}^{T}$. If the forecast x_t^{t+1} contains pertinent information of y_{t+1} for $t = 1, \cdots, T$, including it in D_t will increase the precision of μ_t for $t = 1, \cdots, T$ and reduce Σ_{uu}, the variance of $\{u_{t+1}\}_{t=1}^{T}$. Hence, the ratio of Σ_{uu} derived under (4.10) and (4.7) is regarded as a measure of the information content in the current forecast $\{x_t^{t+1}\}_{t=1}^{T}$. When estimating the predictive systems via the Gibbs sampler, the same random number generator seeds are used for different predictive systems to eliminate the effects of different random numbers.

4.3.3 Measuring the information content of competing forecasts

For the case when researchers have competing forecasts, the predictive system can be separately applied to the competing forecasts. For example, if there are two sets of competing forecasts, x_{1t}^{t+1} and x_{2t}^{t+1} for y_{t+1} for $t=1,\cdots,T$, the following two equations can be separately combined with equations (4.6) & (4.8) to form two predictive systems.

$$x_{1t+1}^{t+2}=\theta+Ax_{1t}^{t+1}+v_{t+1} \tag{4.11}$$

$$x_{2t+1}^{t+2}=\theta+Ax_{2t}^{t+1}+v_{t+1} \tag{4.12}$$

In the first predictive system consisting of equations (4.6), (4.8), and (4.11), $D_t=\{y_t,y_{t-1},\cdots,x_{1t}^{t+1},x_{1t-1}^{t},\cdots\}$. In the second predictive system consisting of equations (4.6), (4.8), and (4.12), $D_t=\{y_t,y_{t-1},\cdots,x_{2t}^{t+1},x_{2t-1}^{t},\cdots\}$. If x_{1t}^{t+1} contains more information related to y_{t+1} than x_{2t}^{t+1} for $t=1,\cdots,T$, the first predictive system's estimates of $\{\mu_t\}_{t=1}^{T}$ will be more accurate than those of the second predictive system. Consequently, the first predictive system's estimate of Σ_{uu} will be smaller than that of the second predictive system. This reduction in the variance of the un-forecasted shocks is regarded as the relative measure of the information content of $\{x_{1t}^{t+1}\}_{t=1}^{T}$ over $\{x_{2t}^{t+1}\}_{t=1}^{T}$. Essentially, forecasts' information contents are determined by the size of the predictive system's un-forecasted part of the variable being forecasted. The greater the ability of the predictive system to estimate the states of y_t , the greater the information contained in the variables in its information set.

4.3.4 Forecasts of other horizons

The above two sections only discuss the information content of the next period's forecast. Although forecasts for the immediate future may be the most useful and sought after, quite often forecasts for various other horizons are available. In the present framework, it is straightforward to measure the information content of these forecasts. For example, for the 2-period-ahead forecast, x_t^{t+2} , equation (4.7) can be replaced by,

$$x_t^{t+2}=\theta+Ax_{t-1}^{t+1}+v_{t+1}.$$

Therefore, $D_t = \{y_t, y_{t-1,} \cdots, x_{t-1}^{t+1}, x_{t-2}^{t}, \cdots\}$. Since x_{t-1}^{t+1} is the 2-period-ahead forecast of y_{t+1} for $t = 1, \cdots, T$, if x_{t-1}^{t+1} contains useful information about y_{t+1}, the precision of $\{\mu_t\}_{t=1}^{T}$ will be increased. The above measures of information contents could be used. In general, for the n-period-ahead forecast, x_t^{t+n}, equation (4.7) can be substituted by,

$$x_{t+2-n}^{t+2} = \theta + Ax_{t+1-n}^{t+1} + v_{t+1}.$$

To measure the information content of a single set of forecasts, similar modifications can be made to equation (4.10)

4.3.5 The advantages of the proposed measures

From the above discussion, the proposed measures of information content have the following advantages over the regression-based measures. First, the proposed measures do not suffer from the multicollinearity problem that affected the regression-based measures as explained before. Second, the proposed measures are derived from well-defined information sets. It is straightforward to control for the desired conditioning information, and to see where the differences in information contents come from. Finally, the regression approach in equations (4.1) and (4.2) is too restrictive in modeling forecasts $\{x_t^{t+1}\}_{t=1}^{T}$ as an exact linear function of the actual value $\{y_{t+1}\}_{t=1}^{T}$ (or the unobserved states $\{\mu_t\}_{t=1}^{T}$). It seems more likely that the forecasts are imperfect, in that they are correlated with $\{y_{t+1}\}_{t=1}^{T}$ but cannot deliver them perfectly. For example, there may be periods when the variable being forecasted is quite stable and periods when it is very unstable. As a result, the forecasting difficulty will vary accordingly. Therefore, there will be periods when the correlation between forecasts and actual values is high and periods when the correlation is low. The predictive system regards forecasts as relevant variables that could help infer the unobserved states $\{\mu_t\}_{t=1}^{T}$. The predictive system uses the information in the forecasts through their innovations, without imposing an exact linear relationship between forecasts and actual values throughout the sample.

4.4 Empirical Applications

4.4.1 Data

To compare with earlier studies, Romer and Romer (2000)'s data set is chosen to calculate the proposed measures of information content. The present study focuses on the comparison of Federal Reserve's Green Book (GB) and Survey of Professional Forecasters' (SPF) inflation (GNP deflator) forecasts. Before each meeting of Federal Open Market Committee (FOMC), the Federal Reserve staff prepares forecasts of key economic variables which are presented in the GB. Since the information content measures proposed in the present study require the data to be of uniform frequency and FOMC meetings take place roughly every six weeks, quarterly series of GB forecasts are constructed by taking the FOMC meeting date closest to the middle of each quarter. The SPF survey begun in 1968 as a project of the American Statistical Association and the National Bureau of Economic Research, and was taken over in 1990 by the Federal Reserve Bank of Philadelphia. The SPF forecasts are quarterly and the mean of all survey participants' forecasts is taken to be the consensus forecast.

Because of Federal Reserve's policy of releasing its forecasts with a five-year lag, the above two sets of inflation forecasts are ideal for the comparison of information content. Since SPF forecasters do not have real time access to Federal Reserve's forecasts, they can not infer Federal Reserve's private information content by studying its forecasts. The sample period is from the 4^{th} quarter of 1968 to the 2nd quarter of 1991, for a total of 91 observations. There are some missing values of the 2- to 4-quarter-ahead forecasts at the beginning of the sample. Because SPF seldom makes forecast more than 4 quarters ahead, the present study analyzes current quarter forecasts to 4-quarter-ahead forecasts. Following Romer and Romer (2000), the second revisions of actual data are used. The rationale for proceeding in such way is that the forecasting literature usually treats near-term revisions as actual values.

Table 4.1 shows the correlation matrix of SPF forecasts, GB forecasts, and actual values. Two stylized facts can be observed in Table 4.1. First, the SPF and GB forecasts are more correlated between themselves than they are with actual values. This legitimizes the collinearity concern in the introduction. Because they are highly correlated, SPF and GB forecasts in some sense have a lot of common information content. Second, GB forecasts are slightly more correlated with actual values than SPF forecasts. Based on the correlation coefficients, one would expect the GB forecasts to be slightly better than the SPF forecasts.

Table 4.1 The Correlations of Inflation Forecasts and Actual Values

Current Quarter			
	Actual	GB	SPF
Actual	1	0.87	0.84
GB		1	0.93
SPF			1

1 Quarter Ahead			
	Actual	GB	SPF
Actual	1	0.79	0.70
GB		1	0.94
SPF			1

2 Quarters Ahead			
	Actual	GB	SPF
Actual	1	0.69	0.56
GB		1	0.91
SPF			1

3 Quarters Ahead			
	Actual	GB	SPF
Actual	1	0.65	0.43
GB		1	0.88
SPF			1

4 Quarters Ahead			
	Actual	GB	SPF
Actual	1	0.77	0.59
GB		1	0.89
SPF			1

4.4.2 Replication of the basic model in Romer and Romer (2000)

To facilitate comparison of the present results with those of previous studies, the parameters of the basic model in Romer and Romer (2000) (equation 2) are estimated using the present data and presented in Table 4.2. The results in Table 4.2 are very similiar to those in Romer and Romer (2000)'s study. The minor differences are probably due to slight differences in the samples, as the data set used here contains more observations. The point estimates of GB forecasts are typically between one and two. The point estimates of SPF forecasts are typically smaller than zero. Romer and Romer (2000) interpreted the results as the Federal Reserve possessing valuable information not contained in the SPF forecasts. According to their interpretation, GB forecasts dominate SPF forecasts for all forecast horizons, since the coefficients corresponding to SPF forecasts are significantly less or equal to zero, and the coefficients associated with GB forecasts are significantly larger than zero.

Table 4.2 Estimates of the Basic Model in Romer and Romer (2000)

Forecast Horizon	α	β (SPF)	γ (GB)	R^2	N
Current Quarter	0.02 (0.37)	0.19 (0.18)	0.82 (0.16)	0.76	91
1-Quarter-Ahead	0.42 (0.49)	-0.42 (0.26)	1.39 (0.24)	0.62	91
2-Quarter-Ahead	1.15 (0.64)	-0.65 (0.29)	1.51 (0.26)	0.49	90
3-Quarter-Ahead	1.96 (0.70)	-1.09 (0.27)	1.83 (0.24)	0.51	87
4-Quarter-Ahead	-0.08 (0.67)	-0.63 (0.24)	1.66 (0.24)	0.62	64

Notes:

1. The regression specification is $y_{t+s} = \alpha + \beta x_{tSPF}^{t+s} + \gamma x_{tGB}^{t+s} + \varepsilon_{t+s}$, where y_{t+s} denotes the actual inflation, x_{tSPF}^{t+s} denotes SPF inflation forecasts, x_{tGB}^{t+s} denotes GB inflation forecasts. s and t index the horizon and date of the forecasts.

2. Standard errors are in the parentheses.

3. N is the number of observations.

4.4.3 Comparing information content of GB forecasts with SPF forecasts

Using the predictive system, the variances of un-forecasted shocks of the GB and SPF forecasts are separately estimated, and denoted as Σ_{uu}^{GB} and Σ_{uu}^{SPF} respectively. The variance ratio, $\Sigma_{uu}^{GB} / \Sigma_{uu}^{SPF}$, is then calculated. Table 4.3 gives the posterior means and standard deviations of the variance ratios as well as the percentages of the draws that are less than one. It shows that GB forecasts contain more information about actual values than SPF forecasts for the current quarter and 1-quarter-ahead forecasting. But for more than 1-quarter-ahead forecasts, GB's information advantage is statistically insignificant.

The GB's advantage over SPF is most noticeable for current quarter forecasts. The posterior mean of $\Sigma_{uu}^{GB} / \Sigma_{uu}^{SPF}$ is 0.73 for the current quarter forecasts. About 99.9% of the $\Sigma_{uu}^{GB} / \Sigma_{uu}^{SPF}$ draws are less than one for the current quarter forecasts. These two numbers drop to 0.86 and 97.27% respectively for the 1-quarter-ahead forecasts. The posterior means of $\Sigma_{uu}^{GB} / \Sigma_{uu}^{SPF}$ for 2- to 4-quarter-ahead forecasts are not significantly less than one, as the percentages of $\Sigma_{uu}^{GB} / \Sigma_{uu}^{SPF}$ draws smaller than one are all less than 95% for these three sets of forecasts. For the 4-quarter-ahead forecasts, the posterior mean of $\Sigma_{uu}^{GB} / \Sigma_{uu}^{SPF}$ is 0.98, and only 69.63% of the $\Sigma_{uu}^{GB} / \Sigma_{uu}^{SPF}$ draws are less than one. The histograms of $\Sigma_{uu}^{GB} / \Sigma_{uu}^{SPF}$ are shown in Figure 4.1. As expected from Table 4.3, for current quarter and 1-quarter-ahead forecasts, most of the mass of the distributions are to the right of a variance ratio equal to one. As the forecast horizon gets longer, the distribution gets more dispersed and it has more mass for ratios greater than one.

An interesting fact about Table 4.3 and Figure 4.1 is that the decrease in GB forecasts' information content over SPF is not monotonic. The 3-quarter-ahead forecasts are better than 2-quarter-ahead forecasts. This may be due to seasonality in macroeconomic series. At time t, last quarter ($t-1$)'s actual values of most macroeconomic series are known. If there is seasonality in quarterly inflation rates and in the series that forecasters use to forecast inflation, the actual value of inflation rate and its related series at $t-1$ may have relatively more information about the inflation rate at $t+3$ than about the inflation rate at $t+2$, since $t-1$ and $t+3$ are exactly 4 quarters apart. Therefore, GB could forecast $t+3$ inflation rate relatively better than $t+2$ at time t. Since forecasters probably draw from many macroeconomic series to forecast inflation rates and it is not feasible to determine exactly which related variables GB and SPF use in making their forecasts, it is not possible to formally test the above hypothesis. However, the fact that most macroeconomic series have a seasonal component gives comfort to the explanations above.

The above results show that GB forecasts do contain more information about the actual values than SPF forecasts. But this advantage is only restricted to short term forecasting. As forecast horizons get longer, the informational advantage becomes statistically insignificant. Intuitively, this makes sense, since it is very difficult to accurately forecast macroeconomic shocks multiple periods into the future. Hence, the decline of the informational advantages, if any, should be expected. In contrast, Romer and Romer (2000) concluded that GB forecasts have additional information for all the forecast horizons. The limitation of their approach is that the standard errors in the OLS regressions are too large to distinguish one regression from another.

Table 4.3 Descriptive Statistics of $\Sigma_{uu}^{GB} / \Sigma_{uu}^{SPF}$

Forecast horizon (Quarters)	Posterior Mean	Posterior Standard Deviation	Percentage Ratio<1
0	0.73	0.07	99.90%
1	0.86	0.06	97.27%
2	0.95	0.05	87.67%
3	0.87	0.09	91.67%
4	0.98	0.05	69.63%

Note: The variance ratio refers to the ratio of Σ_{uu} s derived separately using GB and SPF forecasts by the predictive system. A ratio less than 1 indicates that GB forecasts have additional information not contained in SPF forecasts.

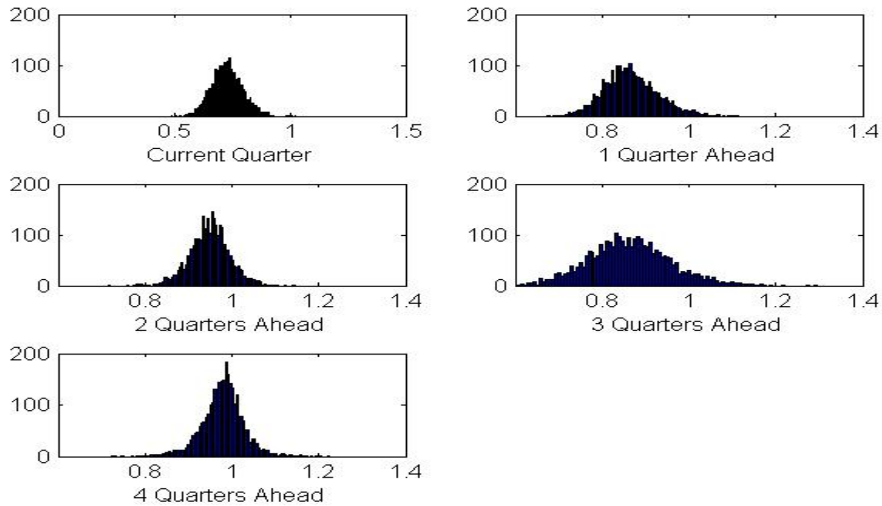


Figure 4.1 Histograms of $\Sigma_{uu}^{GB} / \Sigma_{uu}^{SPF}$

4.4.4 Information contents of GB and SPF forecasts

Subsection 4.3.2 outlines the measure to determine the information contents of one set of forecasts. This section shows how to estimate and use this measure by studying GB and SPF forecasts separately. Denote Σ_{uu}^{C} as the variance of un-forecasted shocks in the predictive system consisting of equations (4.6), (4.8), and (4.7). Denote Σ_{uu}^{Lag} as the variance of un-forecasted shocks in the predictive system consisting of equations (4.6), (4.8), and (4.10). Table 4.4 gives the posterior means and standard deviations of $\Sigma_{uu}^{C} / \Sigma_{uu}^{Lag}$ as well as the percentages of the $\Sigma_{uu}^{C} / \Sigma_{uu}^{Lag}$ draws that are smaller than one for GB and SPF forecasts.

Panel A in Table 4.4 shows that if current GB forecasts are included in the calculation of μ_t, Σ_{uu} could be reduced by 50% and 25% for the current quarter and 1-quarter-ahead prediction, respectively. This shows that GB short term forecasts do have significant additional information compared to forecasts issued earlier. But for longer horizon forecasts, the reduction in Σ_{uu} is not significant. For 2- and 4-quarter-ahead forecasts, none of the posterior means of $\Sigma_{uu}^{C} / \Sigma_{uu}^{Lag}$ are significantly less than one. For the 3-quarter-ahead forecasts, 95.7% of $\Sigma_{uu}^{C} / \Sigma_{uu}^{Lag}$ draws are smaller than one. Figure 4.2 depicts the histograms of $\Sigma_{uu}^{C} / \Sigma_{uu}^{Lag}$ for the GB forecasts. It shows overwhelmingly favorable evidence for the current quarter forecasts as the $\Sigma_{uu}^{C} / \Sigma_{uu}^{Lag}$ mass is smaller than one. But for other forecast horizons, the $\Sigma_{uu}^{C} / \Sigma_{uu}^{Lag}$ masses are generally centered around one. These results are suggestive evidence that GB long term forecasts do not contain much information about actual values.

Panel B in Table 4.4 shows that if current SPF forecasts are included in the calculation of μ_t, Σ_{uu} could be reduced by 20% for the current quarter forecasting. Surprisingly, the reduction in Σ_{uu} is minimal for 1- to 4-quarters ahead forecasts. As a matter of fact, the posterior means of $\Sigma_{uu}^{C} / \Sigma_{uu}^{Lag}$ are larger than one for 2- to 4-quarter-ahead forecasts. This indicates that SPF forecasts contain very little information about actual values except for current quarter forecasts. Figure 4.3 depicts the histograms of $\Sigma_{uu}^{C} / \Sigma_{uu}^{Lag}$ for the SPF forecasts. It is obvious that the majority of the mass correspond to $\Sigma_{uu}^{C} / \Sigma_{uu}^{Lag}$ ratio is larger than one for the 2- to 4-quarter-ahead forecasts. Including the SPF's 2- to 4-quarter-ahead forecasts in the predictive system only adds noise to the estimate of $\{\mu_t\}_{t=1}^{T}$. In other words, nothing will be lost by using only past inflation values in the predictive system, i.e., using equations (4.6) and (4.8) alone.

Table 4.4 Descriptive Statistics of $\Sigma_{uu}^{C} / \Sigma_{uu}^{Lag}$

Panel A: Federal Reserve Green Book Forecasts

Forecast horizon (Quarters)	Posterior Mean	Posterior Standard Deviation	Percentage (Ratio<1)
0	0.50	0.09	100.00%
1	0.75	0.09	99.40%
2	0.94	0.05	88.70%
3	0.79	0.12	95.70%
4	0.94	0.05	87.10%

Panel B: Survey of Professional Forecasters Forecasts

Forecast horizon (Quarters)	Posterior Mean	Posterior Standard Deviation	Percentage (Ratio<1)
0	0.80	0.07	98.65%
1	0.97	0.06	73.10%
2	1.03	0.03	17.93%
3	1.02	0.04	27.47%
4	0.98	0.05	65.83%

Note: The ratio $\Sigma_{uu}^{C} / \Sigma_{uu}^{Lag}$ refers to the ratio of Σ_{uu} s derived separately using current and lagged forecasts by the predictive system. A ratio less than one indicates that forecasts have additional information not contained in past actual values.

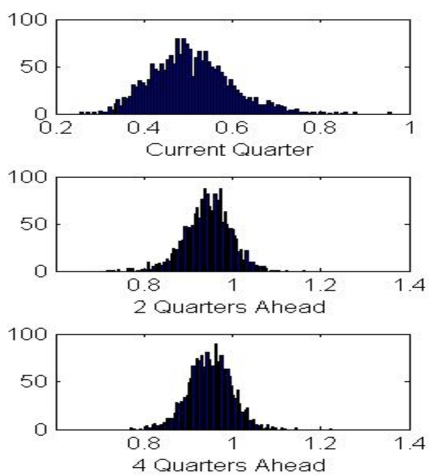


Figure 4.2 Histograms of $\Sigma_{uu}^{C} / \Sigma_{uu}^{Lag}$ of GB Forecasts

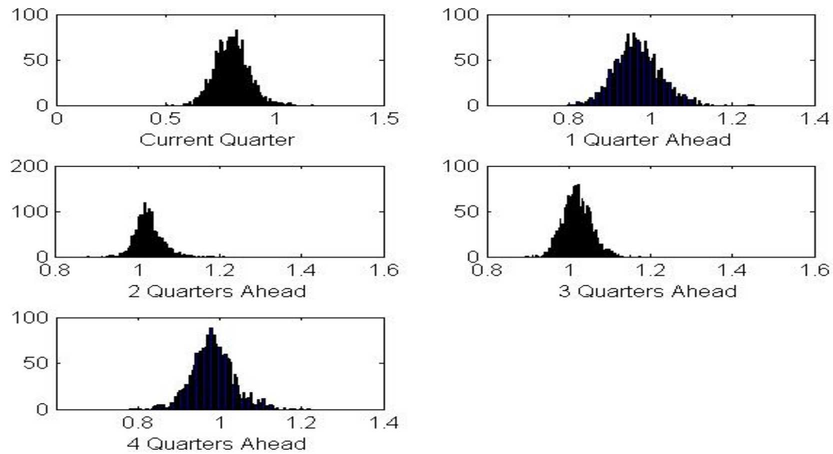


Figure 4.3 Histograms of $\Sigma_{uu}^{C} / \Sigma_{uu}^{Lag}$ of SPF Forecasts

In sum, GB's current and 1-quarter-ahead forecasts are found to have more information than corresponding SPF's forecasts. GB's 2- to 4-quarter-ahead forecasts and SPF's 1- to 4-quarter-ahead forecasts are found to have very little information about the actual values in addition to previous forecasts. The standard regression method would find that GB forecasts contain useful information for all forecast horizons as shown in Table 4.2.

The method proposed in the present study could distinguish the qualities of seemingly identical forecasts. For the 1-quarter-ahead forecasts, the correlation between GB and SPF forecasts is 0.94 and both series are very close to each other as shown in Figure 4.4. However, the proposed measures show that GB forecasts contain extra information and SPF forecasts do not.

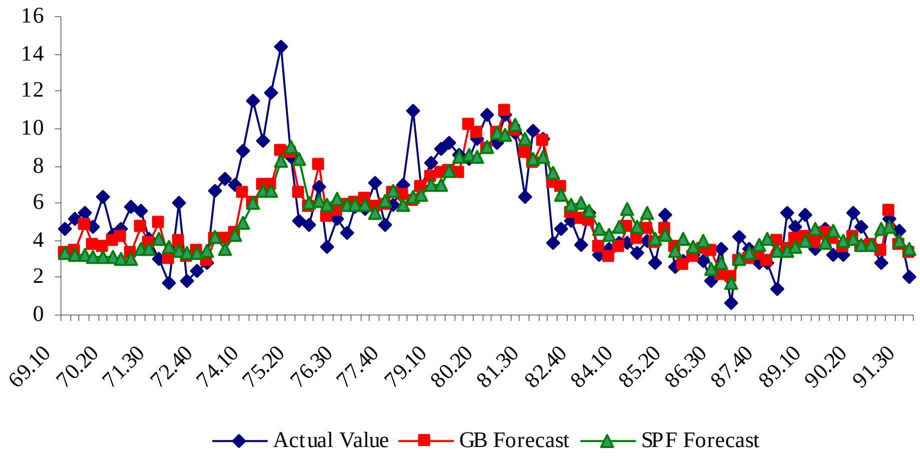


Figure 4.4 Time Series Plot of 1-Quarter-Ahead Forecasts and Actual Values

4.5 Conclusion

The present study proposes a new approach to compare survey forecasts' information content. It differs from the regression-based approach by Fair and Shiller (1989, 1990) which often suffers from multicollinearity problems and makes it difficult to define the conditioning information set. Based on recent developments in the forecasting literature (Pastor and Stambaugh (2006)), the present study regards forecasts as predictors and derives μ_t as the next period's expected values of y_{t+1}, the variable being forecasted. It focuses on the variance of $\{u_{t+1}\}_{t=1}^{T}$, the un-forecasted shocks of $\{y_{t+1}\}_{t=1}^{T}$ conditional on well-defined information sets specified by the researcher. Depending on the forecast quality, the accuracy of $\{\mu_t\}_{t=1}^{T}$ will change. The more accurate $\{\mu_t\}_{t=1}^{T}$, the smaller the variance of $\{u_{t+1}\}_{t=1}^{T}$. The variances of u_{t+1} associated with competing forecasts can be calculated. The ratio of these variances can be regarded as the measure of the relative information contents of the competing forecasts. The present study also proposes a way to measure the information content of forecasts from the same source, without reference to competing forecasts.

The above two measures are applied to the data used in Romer and Romer (2000). The empirical application shows that both measures perform fairly well. The present study is able to find that GB forecasts do dominate SPF forecasts but only in the short term. GB's 2- to 4-quarter-ahead forecasts and SPF's 1- to 4-quarter-ahead forecasts are found to contain very little information about actual values in addition to past inflation values. These results contrast Romer and Romer (2000), who find that GB forecasts dominate SPF forecasts for all horizons. The differences come from the fact that the regression-based method yields large standard errors, which makes the regressions indistinguishable from each other.

The proposed information content measures do not suffer from the multicollinearity problem that could affect previous regression-based measures. The flexibility of the advocated measures allows one to determine the information content of a single set of forecasts, which is not achievable in the regression-based method. The proposed measures are derived from well-defined information sets. The predictive system used in the present study uses the information in the forecasts through their innovations without imposing an exact linear relationship between forecasts and actual values throughout the sample, which is an implicit assumption in the regression-based methods. It is more likely that forecasts are imperfect, in that they are correlated with the actual values but cannot predict them perfectly.

4.6 References

Ang, Andrew, Geert Bekaert and Min Wei (2005), Do macro variables, asset markets or surveys forecast inflation better? NBER Working Paper No. 11538.

Carter, Chris K., Robert Kohn (1994), On Gibbs sampling for state space models, Biometrika, 81, 541-553.

Clements, Michael P., David F. Hendry (1993), On the limitations of comparing mean square forecast errors, Journal of Forecasting, 12, 8, 617-637.

Ericsson, Neil. R. (1993), On the limitations of comparing mean squared forecast errors: clarifications and extensions, Journal of Forecasting, 12, 8, 644-651.

Fair, Ray C., Robert J. Shiller (1990), Comparing information in forecasts from econometric models, American Economic Review, 80, 3. 375-389.

Fair, Ray C., Robert J. Shiller (1989), The informational content of ex ante forecasts, Review of Economics and Statistics, 71, 2. 325-331.

Fruhwirth-Schnatter, Sylvia (1994), Data augmentation and dynamic linear models, Journal of Time Series Analysis, 15, 183-202.

Kalman, Rudolph Emil (1960), A new approach to linear filtering and prediction problems, Transactions of the ASME—Journal of Basic Engineering, 82, Series D, 35—45.

Pastor, Lubos, Robert F. Stambaugh (2006), Predictive systems: living with imperfect predictors. CRSP Working Paper No. 617.

Romer, Christina D., David H. Romer (2000), Federal reserve information and the behavior of interest rates, American Economic Review, 90, 3. 429-457.

Stock, J. H., M. W. Watson (1999), Forecasting inflation, Journal of Monetary Economics, 44, 293-335.

Yock Y. Chong, David F. Hendry (1986), Econometric evaluation of linear macroeconomic models, Review of Economic Studies, 53, 4, 671-690.

4.7 Appendix

This Appendix outline the estimation steps of μ_t, as well as the parameters in the predictive system. The predictive system consists of:

$$y_{t+1} = \mu_t + u_{t+1}$$

$$x_{t+1}^{t+2} = \theta + Ax_t^{t+1} + v_{t+1}$$

$$\mu_{t+1} = \alpha + B\mu_t + w_{t+1}$$

The residuals in the system are assumed to be distributed identically and independently across t as,

$$\begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix} \sim N(0,\Sigma) = N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv} & \Sigma_{uw} \\ \Sigma_{vu} & \Sigma_{vv} & \Sigma_{vw} \\ \Sigma_{wu} & \Sigma_{wv} & \Sigma_{ww} \end{bmatrix} \right).$$

Let D_0 denote the null information set, the unconditional moments are given as:

$$\begin{bmatrix} y_t \\ x_t^{t+1} \\ \mu_t \end{bmatrix} | D_0 \sim N(E,V) = N\left(\begin{bmatrix} E_y \\ E_x \\ E_y \end{bmatrix}, \begin{bmatrix} V_{yy} & V_{yx} & V_{y\mu} \\ V_{xy} & V_{xx} & V_{x\mu} \\ V_{\mu y} & V_{\mu x} & V_{\mu\mu} \end{bmatrix} \right)$$

Let z_t denote the vector of the observed data at time t, $z_t = \begin{bmatrix} y_t \\ x_t^{t+1} \end{bmatrix}$. Therefore, data observed through time t consists of $D_t = (z_1, \cdots, z_t)$, and the complete data is D_T. Also define: $E_z = \begin{bmatrix} E_y \\ E_x \end{bmatrix}, V_{zz} = \begin{bmatrix} V_{yy} & V_{yx} \\ V_{xy} & V_{xx} \end{bmatrix}, V_{zu} = \begin{bmatrix} V_{yu} \\ V_{xu} \end{bmatrix}$. The unconditional moments can be solved as:

$$E_y = \alpha/(1-B),\ E_x = \theta/(1-A)$$

$$V_{\mu\mu} = \Sigma_{ww}/(1-B^2),\ V_{yy} = V_{\mu\mu} + \Sigma_{uu},\ V_{xx} = \Sigma_{vv}/(1-A^2)$$

$$V_{\mu x} = \Sigma_{wv}/(1-BA),\ V_{\mu y} = BV_{\mu\mu} + \Sigma_{wu},\ V_{yx} = AV_{\mu x} + \Sigma_{uv}.$$

Drawing the time series of μ_t

To draw the time series of the unobservable values of μ_t conditional on the current parameter draws, the forward filtering, backward sampling approach, originally developed by Carter and Kohn (1994) and Fruhwirth-Schnatter (1994) is applied.

Filtering

The first stage follows the standard methodology of Kalman filtering. Let Γ denote the parameters in the model. Define,

$$a_t = E(\mu_t | \Gamma, D_{t-1}),\ P_t = Var(\mu_t | \Gamma, D_{t-1})$$

$$b_t = E(\mu_t | \Gamma, D_t),\quad Q_t = Var(\mu_t | \Gamma, D_t)$$

$$f_t = E(z_t \mid \Gamma, D_{t-1}), \qquad S_t = Var(z_t \mid \Gamma, D_{t-1})$$

$$e_t = E(z_t \mid \mu_t, \Gamma, D_{t-1}), \quad R_t = Var(z_t \mid \mu_t, \Gamma, D_{t-1})$$

$$G_t = Cov(z_t, \mu_t \mid D_{t-1})$$

Note that $a_1 = E_y$, $P_1 = V_{\mu\mu}$, $f_1 = E_z$, $S_1 = V_{zz}$, $G_1 = V_{z\mu}$, $\mu_1 \mid D_0 \sim N(a_1, P_1)$, $z_1 \mid D_0 \sim N(f_1, S_1)$, and that $z_1 \mid \mu_1, D_0 \sim N(e_1, R_1)$, where $e_1 = f_1 + G_1 P_1^{-1}(\mu_1 - a_1)$, $R_1 = S_1 - G_1 P_1^{-1} G_1^{'}$. Combining this density with $\mu_1 \mid D_0 \sim N(a_1, P_1)$ gives $\mu_1 \mid D_1 \sim N(b_1, Q_1)$, where $b_1 = a_1 + P_1(P_1 + G_1^{'} R_1^{-1} G_1)^{-1} G_1^{'} R_1^{-1}(z_1 - f_1)$, $Q_1 = P_1(P_1 + G_1^{'} R_1^{-1} G_1)^{-1} P_1$. Continuing in this fashion, all conditional densities could be found to be normally distributed, and the moments for $t = 2, \cdots, T$ are:

$$a_t = \alpha + B b_{t-1}$$
$$P_t = B Q_{t-1} B^{'} + \Sigma_{ww}$$

$$f_t = \begin{bmatrix} b_{t-1} \\ \theta + A x_{t-1}^{t} \end{bmatrix}$$

$$S_t = \begin{bmatrix} Q_{t-1} + \Sigma_{uu} & \Sigma_{uv} \\ \Sigma_{vu} & \Sigma_{vv} \end{bmatrix}$$

$$G_t = \begin{bmatrix} Q_{t-1} B^{'} + \Sigma_{uw} \\ \Sigma_{vw} \end{bmatrix}$$

$$e_t = f_t + G_t P_t^{-1}(\mu_t - a_t)$$

$$R_t = S_t - G_t P_t^{-1} G_t^{'}$$

$$b_t = a_t + P_t(P_t + G_t^{'} R_t^{-1} G_t)^{-1} G_t^{'} R_t^{-1}(z_t - f_t)$$

$$Q_t = P_t(P_t + G_t^{'} R_t^{-1} G_t)^{-1} P_t$$

The values of $\{a_t, b_t, Q_t, P_t\}$ for $t = 1, \cdots, T$ are used for the sampling stage.

Sampling

Let $\zeta_t = \begin{bmatrix} y_t \\ x_t^{t+1} \\ \mu_t \end{bmatrix}$. First sample μ_T from $p(\mu_T \mid D_T)$, the normal density obtained in the last step of the filtering. Then for $t = T-1, T-2, \cdots, 1$, sample μ_T from the conditional density of $p(\zeta_t \mid \zeta_{t+1}, D_t, \Gamma)$. Note that the first two elements of ζ_t, z_t, are already observed and thus need not be sampled. To obtain the conditional density, first note that,

$$\zeta_{t+1} \mid \Gamma, D_t \sim N\left(\begin{bmatrix} b_t \\ \theta + Ax_t^{t+1} \\ a_{t+1} \end{bmatrix}, \begin{bmatrix} Q_t + \Sigma_{uu} & \Sigma_{uv} & Q_t B' + \Sigma_{uw} \\ \Sigma_{vu} & \Sigma_{vv} & \Sigma_{vw} \\ BQ_t + \Sigma_{wu} & \Sigma_{wv} & P_{t+1} \end{bmatrix}\right),$$

$$\zeta_t \mid \Gamma, D_t \sim N\left(\begin{bmatrix} y_t \\ x_t^{t+1} \\ b_t \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_t \end{bmatrix}\right),$$

$$\operatorname{cov}(\zeta_t, \zeta_{t+1}' \mid \Gamma, D_t) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Q_t & 0 & Q_t B' \end{bmatrix}.$$

Therefore, $\zeta_t \mid \zeta_{t+1}, \Gamma, D_t) \sim N(h_t, H_t)$, where,

$$h_t = \begin{bmatrix} y_t \\ x_t^{t+1} \\ b_t \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Q_t & 0 & Q_t B' \end{bmatrix} \begin{bmatrix} Q_t + \Sigma_{uu} & \Sigma_{uv} & Q_t B' + \Sigma_{uw} \\ \Sigma_{vu} & \Sigma_{vv} & \Sigma_{vw} \\ BQ_t + \Sigma_{wu} & \Sigma_{wv} & P_{t+1} \end{bmatrix}^{-1} \begin{bmatrix} y_{t+1} - b_t \\ x_{t+1}^{t+2} - \theta - Ax_t^{t+1} \\ \mu_{t+1} - a_{t+1} \end{bmatrix} \text{ and}$$

$$H_t = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_t \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Q_t & 0 & Q_t B' \end{bmatrix} \begin{bmatrix} Q_t + \Sigma_{uu} & \Sigma_{uv} & Q_t B' + \Sigma_{uw} \\ \Sigma_{vu} & \Sigma_{vv} & \Sigma_{vw} \\ BQ_t + \Sigma_{wu} & \Sigma_{wv} & P_{t+1} \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Q_t & 0 & Q_t B' \end{bmatrix}.$$

The mean and variance of μ_t are taken as the relevant elements of h_t and H_t.

Drawing the parameters

The following describes how to obtain the posterior draws of all parameters conditional on the current draw of the time series of μ_t.

Prior distributions

Let $\beta = \begin{bmatrix} \theta \\ A \\ \alpha \\ B \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv} & \Sigma_{uw} \\ \Sigma_{vu} & \Sigma_{vv} & \Sigma_{vw} \\ \Sigma_{wu} & \Sigma_{wv} & \Sigma_{ww} \end{bmatrix}$. Employing normal prior for β and inverse-Wishart prior for Σ, i.e., $\beta \sim N(\mu_\beta, V_\beta)$ and $\Sigma^{-1} \sim W((\rho\Omega)^{-1}, \rho)$. The values of the hyper-parameter values are given as, $\mu_\beta = 0_4, V_\beta = 1000 * I_4, \Omega = V, \rho = 4$.

Drawing Σ given β

The likelihood function can be expressed as,

$$L(\beta,\Sigma) = \prod_{t=1}^{T} (2\pi)^{-1} |\Sigma|^{-1} \exp(-\frac{1}{2}\sum_{i=1}^{T}(\tilde{y}_t - \tilde{X}_t\beta)'\Sigma^{-1}(\tilde{y}_t - \tilde{X}_t\beta))\text{, where}$$

$$\tilde{y}_t = \begin{bmatrix} y_t \\ x_t^{t+1} \\ \mu_t \end{bmatrix}, \tilde{X}_t = \begin{bmatrix} \mu_{t-1} & 0 & 0 & 0 & 0 \\ 0 & 1 & x_{t-1}^{t} & 0 & 0 \\ 0 & 0 & 0 & 1 & \mu_{t-1} \end{bmatrix}.$$

Combining the likelihood function above with the inverse-Wishart prior for Σ, the following posterior conditional distribution could be attained,

$$\Sigma^{-1} | \bullet \sim W((\rho\Omega + \sum_{i=1}^{T}(\tilde{y}_t - \tilde{X}_t\beta)(\tilde{y}_t - \tilde{X}_t\beta)')^{-1}, T+\rho).$$

Drawing β given Σ

Conditional on μ_t and y_{t+1}, the value of μ_{t+1} is known. Because $\begin{bmatrix} u_{t+1} & v_{t+1} & w_{t+1} \end{bmatrix}' \sim N(0,\Sigma)$, the conditional distribution of $\begin{bmatrix} v_{t+1} & w_{t+1} \end{bmatrix}' | u_{t+1}$ is normal and its mean and covariance could be easily calculated. Denote the conditional mean vector as $MEAN_{vw,t+1}$ and the conditional covariance matrix as COV_{vw}. Subtracting the conditional mean $MEAN_{vw,t+1}$ from $\begin{bmatrix} x_{t+1}^{t+2} & \mu_{t+1} \end{bmatrix}'$, equations (7) and (8) can be rewritten as,

$$\begin{bmatrix} x_{t+1}^{t+2} \\ \mu_{t+1} \end{bmatrix} - MEAN_{vw,t+1} = \begin{bmatrix} 1 & x_t^{t+1} & 0 & 0 \\ 0 & 0 & 1 & \mu_t \end{bmatrix}\beta + \begin{bmatrix} v_{t+1} \\ w_{t+1} \end{bmatrix} - MEAN_{vw,t+1}$$

Let $\begin{bmatrix} \hat{x}_{t+1}^{t+2} \\ \hat{\mu}_{t+1} \end{bmatrix} = \begin{bmatrix} x_{t+1}^{t+2} \\ \mu_{t+1} \end{bmatrix} - MEAN_{vw,t+1}$ and $\begin{bmatrix} \hat{v}_{t+1} \\ \hat{w}_{t+1} \end{bmatrix} = \begin{bmatrix} v_{t+1} \\ w_{t+1} \end{bmatrix} - MEAN_{vw,t+1}$. Stacking series over t and denote $\hat{y} = \begin{bmatrix} \hat{X}_{+1}' & \hat{\mu}_{+1}' \end{bmatrix}', \hat{X} = \begin{bmatrix} 1 & X & 0 & 0 \\ 0 & 0 & 1 & \mu \end{bmatrix}, \hat{e} = \begin{bmatrix} \hat{v}' & \hat{w}' \end{bmatrix}'$. Equations (7) and (8) can be further rewritten as, $\hat{y} = \hat{X}\beta + \hat{e}$, which is a standard Seemingly Unrelated Regression. The posterior conditional for β can be written as,

$$\beta \mid \bullet \sim N(D_\beta d_\beta, D_\beta)\,,$$

where, $D_\beta = (\hat{X}'(COV_{vw} \otimes I_T)\hat{X} + V_\beta^{-1})^{-1}$ and $d_\beta = \hat{X}'(COV_{vw} \otimes I_T)\hat{y} + V_\beta^{-1}\mu_\beta$.

CHAPTER 5. GENERAL CONCLUSION

In this dissertation, I applied Bayesian econometric methods on various aspects of forecast evaluation. The overall objective is to better evaluate forecasts in terms of bias, efficiency, and information content by accounting for the structure of forecasts and directly addressing various critical econometric issues that are ignored by previous studies. Three related studies have been undertaken to address three issues. My first paper studies forecasts' bias and inefficiency after accounting for forecast error correlations. This study offers additional evidence on the irrationality of stock analysts' forecasts and reconciles contradicting results in the previous literature. My second paper studies forecasts' bias and inefficiency after accounting for forecasts' hierarchical structure. It shows that there is heterogeneity in the degrees of analysts' bias and inefficiency. My third paper proposes new measures of forecasts' information content of actual variables. The advocated measures are computed for a well-known data set and yield different, yet compelling, conclusions from those drawn by the previous literature's regression-based measures. Although the three papers in this dissertation studies specific data sets, the employed methods could be easily applied to forecasts with similar structures.

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my thanks to those who helped me with various aspects of conducting research and the writing of this thesis. First and foremost, I am greatly in debt to my major professor Dr. Sergio H. Lence for his valuable guidance, great patience, and generous support throughout my study, research, and the writing of this thesis. He painstakingly and promptly went through numerous drafts of the papers in this dissertation. If not for his great insights, valuable inputs, and words of encouragement, it would take me much longer to finish this dissertation. I would also like to thank my committee members for their efforts and contributions to my dissertation: Dr. Barry L. Falk, Dr. Dermot J. Hayes, Dr. Joseph A. Herriges, Dr. Travis R.A. Sapp, and Dr. Justin L. Tobias. Their valuable comments during my Preliminary Exam and subsequent inputs directed my dissertation to a better direction than I had originally planed. Dr. Justin L. Tobias patiently answered all my questions about Bayesian econometrics. I would also like to thank Dr. Chad E. Hart at Center for Agricultural and Rural Development for various discussions about the third chapter of this dissertation.