Predicting Stock Price by Applying the Residual Income Model and Bayesian Statistics

By

Huong N. Higgins, Ph.D.
Worcester Polytechnic Institute
Department of Management
100 Institute Road
Worcester, MA 01609
Tel: (508) 831-5626
Fax: (508) 831-5720
Email: hhiggins@wpi.edu

and

Flora Lu
Esurance
650 Davis Street
Tel: (415) 875-4656
Fax: (415) 875-4501
Email: flu@esurance.com

September 18, 2006

This paper originates from Flora Lu’s Masters’ thesis at Worcester Polytechnic Institute under the supervision of Huong Higgins and Bal Nandram. The authors acknowledge the contributions of Bal Nandram to the methodologies described in the paper. The authors have benefited from the comments of Sandra Chamberlain, Carlos Morales, and Bharat Sarath and participants at the 2006 New England Statistics Symposium, the 2006 Canadian Academic Accounting Association conference, and the 2006 Annual American Accounting Association conference.
Predicting Stock Price by Applying the Residual Income Model and Bayesian Statistics

Abstract

Over the past decade of accounting and finance research, the Residual Income Model (RIM) has often been applied as a framework for equity valuation. In this paper, we demonstrate methods to improve the implementation of the RIM. Specifically, we use transformations to explore non-linear relationships between price and accounting data, and we use Bayesian statistics to improve the inference mechanics. We show that our proposed methods yield price forecasts that are significantly more accurate than those based on current implementations of the RIM.
Predicting Stock Price by Applying the Residual Income Model and Bayesian Statistics

1. Introduction

The Residual Income Model (RIM) is a widely used framework for equity valuation based on accounting data. In this paper, we demonstrate implementation mechanics using transformations and Bayesian statistics that can improve the accuracy of stock price forecasts based on the RIM. Following Ohlson (1995), we quantify stock price of a single firm as a function of book value, a series of abnormal earnings, and a normally distributed term $v_t$ representing “other information”, or information other than abnormal earnings that is yet to be captured in current financial statements but affects stock price. Other information $v_t$ is crucial and should be thought of as capturing all non-accounting information used in the prediction of future abnormal earnings (Ohlson 1995, Hand 2001). Other information $v_t$ highlights the limitations of transaction-based accounting in determining share prices, because while prices can adjust immediately to new information about the firm’s current and/or future profitability, generally accepted accounting principles (GAAP) primarily capture the value-relevance of new information through transactions (Hand 2001).

Because of the crucial role of $v_t$, the manner in which it is addressed may well determine the empirical success of the RIM. Much empirical research motivated by Ohlson (1995) has set $v_t$ to zero. Because $v_t$ is unspecified, setting it to zero is of pragmatic interest. However, setting $v_t$ to zero would mean that only financial accounting data matter in equity valuation, a patently simplistic view. More recent research has sought to address $v_t$, for example by assuming time series (Dechow et al. 1999, Callen and Morel 2001), and by assuming relations between $v_t$ and
other informational observables (Myers 1999). However, these implementations have met with moderate empirical success, showing that time series structures are no better than book values or dividend-discount models in predicting security prices, and complicated models perform worse than simple ones. The surprising fact that informational variables do not lead to improvement indicates that the benefit of information is offset by weakened inference mechanics due to increased complexity. Overall, the modest success of prior implementations suggests two problems: 1) the true information dynamic underlying $v_t$ might be some other process than described by the RIM, and/or 2) the dynamic is lost in statistical inference.

We seek improvement by addressing these two problems above. Our approach is not by incorporating additional informational variables as prior research has attempted without success. Instead, we address the stochastic process and statistical inference by using transformations and applying an alternative statistical inference method. First, we use transformations to explore alternative processes for the RIM, essentially considering non-linear relationships between valuation and accounting variables. While Ohlson (1995) proposes linearity, there is a fair amount of evidence of non-linearity (Hand 2001, who cites many studies with such evidence). Second, we apply an alternative statistical inference method because the one used in current research has room for improvement. As discussed above, the fact that informational variables do not lead to better forecasts is surprising, because if added variables carry relevant content, an effective inference mechanic should yield better forecasts. Current research has typically relied on maximum likelihood methods for inference. We point out the drawbacks of these methods, and discuss an alternative inference method, Bayesian statistics, that does not have such

---

1 By informational variables, we mean variables adding substantial information carrying fundamental economic content that should explain valuation. Myers (1999) refers to these variables as conditioning variables.
drawbacks. Overall, our contribution is to improve the implementation of the RIM, specifically by applying alternative stochastic processes and inference mechanics for the RIM.

We operationalize the RIM based on abnormal quarterly earnings. We include four quarters of abnormal earnings for each valuation equation, therefore mitigating seasonality patterns. The use of quarterly (versus annual) earnings comes with some practical advantages. First, many investment professionals’ activities are tied to quarterly forecasts. Second, quarterly data allow a reasonable time series based on a short history, which mitigated confounding factors from macro-economic cycles, and allows analysis of young firms. Third, as we find through our analyses, quarterly data are more stationary, mitigating estimation and prediction problems.² Finally, prior accounting research has focused on annual earnings, consequently there is currently little documentation of valuation results based on quarterly earnings. The accuracy of our price estimates based on quarterly data are comparable with prior results based on annual data even before applying transformations and Bayesian statistics.

Because we aim to forecast without hindsight knowledge of actual earnings, abnormal earnings is measured as the difference between analyst forward earnings forecast (best knowledge of actual earnings) and the earnings number achieved under growth of book value at normal discount rate. As we use four quarterly abnormal earnings numbers, all four quarterly earnings forecasts are provided by I/B/E/S. We use the time varying quarterly Treasury bill rate in all our models.³

While we do not seek sophisticate models of discount rates, we are able to highlight improvement

² For example, Myers (1999) discusses that Ordinary Least Square cannot be used for estimating the linear information dynamics of non-stationary variables.
³ We do not seek to model the discount rates. In fact, similar to prior studies that explore informational variables, prior research seeking to model the discount rate has met with poor results. Myers (1999) states that estimations of industry or firm-specific discount rates are severely problematic. Sougiannis and Yaekura (2000) use four alternative discount rates, and find that the more specific rate leads to the worse performance model. Bernard (1995) set the discount rate uniformly at 0.13 in his analysis. Callen and Morel (2001) add a time invariant industry risk-premium to the Treasury bill rate.
from Bayesian mechanisms by applying the discount rates in the same way for both maximum likelihood and Bayesian analyses.

Focusing on SP500 firms, we use 23 quarters of data starting in Q1 1999 to estimate the prediction models, which we then use to predict stock price in Q4 of 2004. We use two types of estimation approaches, maximum likelihood as commonly used in prior research, and Bayesian statistics. We find that our maximum likelihood forecast errors are comparable to prior results. Our Bayesian analyses result in significantly smaller predictive errors than our maximum likelihood analyses. We perform several transformations, however transformations of the maximum likelihood models do not outweigh the usefulness of applying Bayesian statistics. Combinations of transformations and Bayesian statistics result in the best price forecasts.

In sum, we contribute to the literature by improving the implementation of the RIM, but not via sophistication of informational data. Rather, we rely on the parsimonious model described by equation (1), and resort to transformations to explore alternative stochastic processes, and resort to basic laws of probability through Bayesian statistics to improve the inference mechanics. By showing improvement on the back-to-basics model, we demonstrate how to improve the inference mechanism for implementing the RIM. Because our approach helps inferencing, it should also help models with sophisticate informational variables that are developed elsewhere.

The paper proceeds as follows. Section 2 describes our formulation of the RIM. Section 3 discusses the advantage of Bayesian statistics. Section 4 describes our methodology. Section 5 discusses our data and analyses. Section 6 summarizes and concludes the paper.
2. The Residual Income Model

Following Ohlson (1995), we describe the RIM based on three assumptions, the dividend discount model described by Rubinstein (1976), the clean surplus relationship, and an AR(1) autoregressive structure.

The following regression form can be derived from the first two assumptions (See appendix A for this formulation):

\[ y_t = \beta_0 + \beta_1 bv_t + \sum_{k=1}^{n} \beta_{k+1} x_{t+k-1}^a + v_t = x_t' \beta + v_t \]

\[ k = 1, 2, 3, 4, \ldots, n; \ t = 1, \ldots, T. \]

where \( y_t \) denotes the stock price per share at time \( t \), \( bv_t \) is the book value per share at the beginning of time \( t \), \( x_t^a \) represents the abnormal earning at time \( t \), \( \beta = (\beta_0, \ldots, \beta_n)' \) is the vector of intercept and slope coefficients of the predictors, \( x_t' = (1, bv_t, x_{t+1}^a, x_{t+2}^a, x_{t+3}^a, x_{t+4}^a, \ldots)' \) is the vector of intercept and predictors, and \( v_t \) is the other information term.

The third assumption assumes that \( v_t \) has a first order autoregressive structure AR(1), which can be described as:

\[ v_t = \rho v_{t-1} + \epsilon_t \]

\[ \epsilon_t \sim N(0, \sigma^2) \]

where \( \rho \) is the correlation coefficient of time series \( v_t \), \( \epsilon_t \) is the white noise, \( \sigma^2 \) is the variance of the white noise.
Note that Ohlson (1995) describes the third assumption of the RIM as:

\[ x_{i,t+1}^a = \omega x_{i,t}^a + v_{it} + \varepsilon_{it+1} \]

\[ v_{it} = \rho v_{i,t-1} + \varepsilon_{it} \]

where \( \omega \) is the coefficient representing the persistence of abnormal earnings. From Ohlson’s (1995) third assumption, we retain only the structure involving \( v_t \). The difference is because Ohlson (1995) has a focus on abnormal earnings and the issue of earnings persistence, which is favorable for the task of forecasting earnings, but this focus creates an intermediate step for the task of forecasting stock price. In essence, our formulation of the RIM may be viewed as the classical RIM by Peasnell (1982) and the assumption that, over a finite horizon, price is affected by other information \( v_t \) which is governed by an AR(1) autoregressive structure.

3. The Advantage of Bayesian Statistics

From (1) and (2):

\[ v_{t-1} = y_{t-1} - x'_{t-1} \beta, \]

\[ v_t = \rho v_{t-1} + \varepsilon_t = \rho \left( y_{t-1} - x'_{t-1} \beta \right) + \varepsilon_t, \]

\[ y_t = x' \beta + \rho \left( y_{t-1} - x'_{t-1} \beta \right) + \varepsilon_t. \]

Therefore, expressions (1) and (2) can be combined as:

\[ y_t = x' \beta + \rho (y_{t-1} - x'_{t-1} \beta) + \varepsilon_t, t = 2, \ldots, T, \]

\( \varepsilon_t \sim iid N(0, \sigma^2), \quad t = 1, \ldots, T, \) \hspace{1cm} (3)

where \( \beta, \rho \) and \( \sigma^2 \) are unknown parameters.

The task of forecasting one-step ahead can be described as:
\[ y_{T+1} \mid y(T), \mu, \beta, \rho, \sigma^2 \sim \text{Normal} \{ x_{T+1} \beta + \rho(y_T - x_T \beta), \sigma^2 \} , \quad (4) \]

where \( y(T) = (y_1, \ldots, y_T) \).

For inference, a forecaster typically computes maximum likelihood estimates (maximum likelihood) for \( \beta, \rho, \sigma^2 \), denoted by \( \hat{\beta}, \hat{\rho}, \hat{\sigma}^2 \), and the forecasted value of \( y_{T+1} \) can be calculated by plugging in the estimated values of the parameters. For example, for a 95% confidence interval, the forecasted value of \( y_{T+1} \) is

\[ x_{T+1} \hat{\beta} + \hat{\rho}(y_T - x_T \hat{\beta}) \pm 1.96 \frac{\hat{\sigma}}{\sqrt{T}} \quad (5) \]

Essentially, maximum likelihood forecasts are based on parameter estimates.

Maximum likelihood generally works well but it has many drawbacks as discussed by Kennedy (2003) and Zellner (1974). Its key drawback is that estimated parameters are assumed fixed and known with certainty. For the RIM, fixed estimated parameters are only applicable when the firms have fixed investment strategies over time, but firms do vary their investment strategies. The assumption of fixed parameters is problematic given that many studies have shown extremely unstable parameter estimates in asset pricing models (starting with Blume 1971). Another drawback of maximum likelihood is that the distribution of the forecasted value is assumed to be normal even after the estimated parameters are plugged in. Essentially, the maximum likelihood approach under-estimates the variability of estimated parameters, and incorrectly assumes the distribution of the forecasted variable.

On the contrary, in the Bayesian paradigm, model parameters are deemed stochastic, assumed to follow distributions that are not fixed and known with certainty. As a result, Bayesian inferencing
should work better than maximum likelihood (Zellner 1974). Stochastic parameters are more adaptive to dynamical changes in the data, or changes in the data that are caused by varying investment strategies, investment skills, and technical innovations. Overall, the Bayesian approach should yield fitter estimation models and stronger predictive power than the maximum likelihood approach because the Bayesian approach is more adaptive to dynamical changes in the data and does not use assumptions that may not be true.

4. Methodology

We explore the stochastic process for the RIM through transformations, and then use both maximum likelihood and Bayesian statistics to fit the selected stochastic processes for inference. First, we use maximum likelihood approach to build estimation models from the first 23 quarters’ data, to predict stock price in the 24th quarter. Besides the un-transformed model, we use many transformations in combination with AR(1) and AR(2) to seek some better than linear maps for the true but unknown information dynamics. We retain the best performing transformations. Then, we apply Bayesian statistics on the untransformed model and the best performing transformations to yield a second set of predicted values. Finally, we compare the maximum likelihood and Bayesian prediction results to demonstrate that Bayesian predictions are more accurate.

We use standard procedures for our Bayesian analysis (for similar applications of these procedures, see Nandram and Petruccelli 1997, and Ying et al. 2005). Bayesian inference depends on taking prior beliefs (prior probabilities) about the parameters before data are available, modifying them in the light of data via likelihood functions, and arriving at posterior probabilities of the parameters via the formulation that posterior is proportional to prior times likelihood (Bayes’ theorem). Then, the posterior density of the predicted value is determined based on the law of total probability, by averaging the predictive distribution over the posterior distribution of
the parameters. Finally, the predicted value is assessed through extensive sampling of the posterior density of the predicted value.

We express the RIM by rewriting (3) to highlight one advantage of the Bayesian method at $t = 1$, which is useless for the maximum likelihood forecaster because $y_0$ and $x'_0$ are unobserved. When $t = 1$, $y_1 = x'_1 \beta + \rho \left( y_0 - x'_0 \beta \right) + \varepsilon_1$. Let $\mu = \rho \left( y_0 - x'_0 \beta \right)$. Then (3) can be expressed as:

\[
y_1 = x'_1 \beta + \mu + \varepsilon_1, \quad y_t = x'_t \beta + \rho (y_{t-1} - x'_{t-1} \beta) + \varepsilon_t, t = 2, \ldots, T, \tag{6}
\]

where $\beta, \mu, \rho$ and $\sigma^2$ are four unknown parameters, which we also express as $\Omega = (\mu, \beta, \rho, \sigma^2)$ to simplify our notation.

Given the previous $T$ periods’ stock prices $y_1, \ldots, y_T$, the task of forecasting the stock price one-step ahead can be described as:

\[
y_{T+1} \mid y_{(T)}, \mu, \beta, \rho, \sigma^2 \sim \text{Normal} \{ x'_{T+1} \beta + \rho (y_T - x'_T \beta), \sigma^2 \},
\]

where $y_{(T)} = (y_1, \ldots, y_T)$.

First, because we view the parameters as stochastic, we assign prior distributions to them. Because we know little about the true stochastic process underlying the information dynamics,
following Jeffreys (1961), we select priors that are non-informative and are invariant to
transformations. So long as priors are non-informative, their selection does not pose an issue,
because non-informative priors do not change the problem. Non-informative priors are certainly
acceptable in case of stationarity, as we later demonstrate with our data. Specifically, we select
our parameters priors as follows:

\[
\pi(\beta) = N_k(\beta \mid \theta, \Delta_0),
\]
\[
\pi(\mu) = N(\mu \mid \mu_0, \sigma^2_0),
\]
\[
\pi(\rho) = U(\rho \mid -1,1),
\]
\[
\pi(\sigma^2) = \Gamma(\sigma^2 \mid a, b).
\]

(7)

Second, we obtain the posterior density \( \pi(\Omega \mid y_{(T)}) \) by updating the priors using the data. Third,
using the law of total probability, we determine the posterior density of the predicted value by
averaging the predictive distribution over the posterior distribution of the parameters.

\[
p(y_{T+1} \mid y_{(T)}) = \int p(y_{T+1} \mid \Omega, y_{(T)})\pi(\Omega \mid y_{(T)})d\Omega.
\]

(8)

The final step is to sample from \( p(y_{T+1} \mid y_{(T)}) \) to help determine \( y_{T+1} \). If the model is correct,
then all information about \( y_{T+1} \) lies in the posterior density \( p(y_{T+1} \mid y_{(T)}) \). Because the posterior
density is not normal, we use Gibbs sampler to obtain a large enough sample from

\[
p(y_{T+1} \mid y_{(T)}).
\]

Our untransformed model is expressed as:
\[ y_t = \beta_0 + \beta_i bv_t + \sum_{k=1}^{4} \beta_{i+k-1} x_{t+k-1}^a + v_t = x_t^a \beta + v_t, \]
\[ k = 1, 2, 3, 4; t = 1, \ldots, 23. \]

\[ v_t = \rho v_{t-1} + \epsilon_t \]
\[ \epsilon_t \sim \mathcal{N}(0, \sigma^2) \]
\[ x_{t}^a = x_t - r_t bv_{t-1} \]
\[ bv_t = bv_{t-1} * (1 + r_t) \]

The predictors in model (9) parallel analysts’ information in their forecasting task: in the middle of quarter \( t \), analysts’ knowledge consists of book value at the beginning of the quarter (\( bv_t \)) and its prior value (\( bv_{t-1} \)), quarterly earnings forecasts of the current quarter (\( x_t \)), quarterly earnings forecasts of 1, 2, and 3 quarters ahead (\( x_{t+1}, x_{t+2}, x_{t+3} \)), and the current quarterly Treasury bill rate (\( r_t \)).

We perform many transformations of Model (9), seeking alternative parameter distributions that may improve model fit and predictive power. The transformations are log, square root, cubic root, and inverse functions, in combination with a second order autoregressive structure AR(2). All models are described in Table 1.

5. Data and Analyses

In this section, we discuss our sample selection and data, our maximum likelihood and Bayesian analyses, and compare the maximum likelihood with Bayesian results.

5.1. Sample Selection and Data

Sample firms are from the SP500 index as of May 2005. The selection criteria are:
a) Price and book value data must be available continuously for 24 quarters, from Q1 1999 through Q4 2004 (Source: Worldscope and Datastream/Thomson Financial).

b) Quarterly earnings forecasts must be available for the current, and one, two, and three quarters ahead for all quarters (Source: I/B/E/S/Thomson Financial).

c) Book values must be greater than zero in all quarters.

This selection process yields 222 firms, which consists of 172 industrial firms, 44 financial firms, and 6 firms from the Transportation and Utility sectors.

Book value is computed as (total assets - total liabilities - preferred stock)/number of common shares. The number of common shares is adjusted for stock splits and dividends. Book value and price data are retrieved from Worldscope. Treasury bill rates are from Datastream, and earnings forecasts from I/B/E/S.

Table 2 shows selected descriptive statistics for the total sample. The median values for price per share, book per share are $33.01 and $8.68, respectively. The median quarterly forecasts of the current quarter, and one, two, and three quarters ahead are $0.36, $0.38, $0.4, and $0.41, respectively. The median quarterly Treasury bill rate is 1.96%.

5.2. Maximum likelihood models

The PROC AUTOREG procedure from SAS is used to estimate the parameters of all models from Table 1. Firms are pooled by sector (industrial or financial). The models’ goodness of fit is assessed based on R-square, and also on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The results (untabulated) show that the AR(2) structure is no better than AR(1), in line with the conclusion drawn by Callen (2001). Further, the inverse is less fit, but the cubic root, log, and square root transformations are better fit than the untransformed
model. Overall, judging from the three above criteria, cubic and log transformations are the two best transformations that improve models’ fit. Therefore, we only report our results based on the untransformed, cubic and log models with AR(1) structure.

The PROC AUTOREG procedure from SAS is also used to predict price per share in Q4 2004.
To assess the models’ predictive power, we measure prediction error as
$$ R = \left| \frac{\hat{y}_{24} - y_{24}}{y_{24}} \right|, $$
where $R$ is the absolute relative difference of predicted stock price over actual stock price for a company, and $\hat{y}_{24}$ and $y_{24}$ are respectively the predicted and actual price per share of a company for the 24th quarter. $R$ close to zero means low prediction error, or high accuracy.

Table 3 shows the results of untransformed model in Panel A, the log transformation in Panel B, and the cubic root transformation in Panel C. In each panel, we report separately for the industrial group and the financial group. The reported results are estimated model coefficients ($\beta_0$–$\beta_3$), the time series correlation coefficient $\rho$, and the fitness of estimation model (Total R-square).

As shown, the estimated model coefficients $\beta_2$–$\beta_3$ are significantly positive in most models, indicating that abnormal earnings in future quarterly earnings contribute to current valuation. All models have high R-square, ranging from 75.87% (untransformed model of the industrial group) to 90.91% (cubic root transformation of the financial group).4 Both transformed models have higher R-squares than the un-transformed model, although transformations result in only slightly higher R-squares for the financial group. In all models, the time series correlation coefficient $\rho$ is negative, consistent with mean reversion, and with absolute values less than 1, consistent with stationarity in the other information term $v_t$.

---

4 Bernard (1995) reports an average R-square of 68% in his Table 1. DeChow et al. show R-squares of 40% and 69% in their Table 6.
5.3. Bayesian Models

The Bayesian procedures that we apply are standard for Bayesian estimation and prediction, which are gaining increasing popularity in business applications (for example, see Nandram and Petruccelli 1997, and Ying et al. 2005). We set up the Bayesian procedures in five steps, setting the likelihood function, setting the posterior distribution of the parameters, Gibbs Sampling, assessing the model fit, and forecasting. Appendix B provides the details of these procedures.

We pool firms by sector (industrial or financial), and show the results of our Bayesian estimation in Table 4. As before, we report the untransformed, cubic and log models with AR(1) structure in Panels A, B, and C respectively. We assess the models’ goodness of fit based on conditional predictive ordinate (CPO) as discussed in Appendix B. A larger Log(CPO) denotes better fit.

In terms of model fitness, Table 4 shows generally better fit among the transformed than un-transformed models. For example, in the industrial sector, Log(CPO) of the un-transformed model is -4.47, of the log model is -4.26, and of the cubic root model is -3.10. Among the financial sector, cubic root transformation results in better fit than the un-transformed model, but the log transformation is not: Log(CPO) of the un-transformed model is -4.65, of the log model is -4.72, and of the cubic root model is -3.44.

The conditional predictive ordinate (CPO) is a Bayesian diagnostic which detects surprising observations. It is calculated through the conditional posterior distribution function. From Gibbs Sampling, we get sample values for the parameters. The most surprising value has minimum CPO, or the maximum log(CPO). We collect 10,000 different values of log(CPO) through Gibbs Sampling, to use the median value. We compare the median values from different models to decide which one does a better job. A larger median value of log(CPO) denotes better fit.
5.4. Comparing Maximum Likelihood with Bayesian Prediction Results

Table 5 compares the prediction results between Maximum Likelihood and Bayesian methods for the industrial sector and the financial sector. For each group, we report the mean absolute prediction error from both Maximum Likelihood and Bayesian methods, based on the un-transformed, log, and cubic root models.

<Table 5 about here>

In the industrial sector, Bayesian prediction errors are smaller than maximum likelihood errors, with or without transformations. Based on the un-transformed model, the mean maximum likelihood prediction error is 45.84%, significantly larger than the Bayesian prediction error of 18.20% (T-test p-value = 0.0001). Based on the log model, the mean maximum likelihood prediction error is 44.34%, significantly larger than the Bayesian prediction error of 12.22% (T-test p-value = 0.0001). Based on the cubic root model, the maximum likelihood prediction error is 43.37%, significantly larger than the Bayesian prediction error of 11.99% (T-test p-value = 0.0001). Transformations reduce the maximum likelihood error by 1.5% (45.84%-44.34%) and 1.97% (45.84%-43.37%) for log and cubic models, respectively. But the transformations are more beneficial to Bayesian predictions, reducing error by 5.98% (18.20%-12.22%) and 6.21% (18.20%-11.99%), respectively. Results based on medians are similar.

For financial firms, prediction errors are generally smaller than those of industrial firms, probably due to higher homogeneity and additional regulatory disclosure requirements in the financial sector. Similarly to the industrial group, Bayesian prediction errors are smaller than maximum

---

Using the same definition of prediction error, DeChow et al. (1999) report errors ranging from 40.2% to 44.5% in Panel B of their Table 5.
likelihood errors with or without transformations. The untransformed prediction error under maximum likelihood and Bayesian are 16.75% and 9.29% (significantly different at T-test p-value = 0.0050). The log prediction error is 19.91% under maximum likelihood, and 11.62% under Bayesian (significantly different at T-test p-value = 0.0482). Likewise, the cubic root prediction error is significantly different under maximum likelihood and Bayesian (17.40% and 9.60%, respectively, for a T-test p-value = 0.0136).

Overall, Table 5 results demonstrate that indeed, Bayesian prediction results are more accurate than maximum likelihood results.

6. Conclusion

Although the Residual Income Model (RIM) proposed by Ohlson (1995) has been embraced enthusiastically as a theoretical framework for equity valuation, it has not fared well in empirical testing. Many implementations of the model have yielded high valuation errors. The issue is probably not omitted variables, because more complicated models incorporating additional observables fare worse than more parsimonious models. From the limited empirical success reported by prior research, one wonders if the true information dynamic of the RIM follows some alternative stochastic processes that have not been captured, or if it is lost in statistical inference.

This paper demonstrates a way to improve implementation of the RIM by exploring alternative stochastic processes underlying the model’s information dynamic using transformations, and by fitting the selected processes with Bayesian statistics for improved inference. We argue that the classical maximum likelihood method underestimates variability in the estimated parameters and incorrectly assumes normal distribution in the predicted dependent variable. Because maximum likelihood assumptions may not be true, this inference method leaves room for improvement. We use transformations to explore some better than linear maps for the true but unknown information
dynamics. Then we use Bayesian processes to fit the selected processes for inference. Unlike in maximum likelihood, the Bayesian approach views model parameters as following distributions that are not fixed and known with certainty. Because it does not use assumptions that may not be true, Bayesian inference is expected to yield better fit in the estimation model, and higher prediction accuracy.

We document the following incidental findings. The AR(2) structure does not provide much improvement over AR(1), consistent with Callen and Morel (2001). Transformations, particularly log and cubic root transformations, yield better estimation fit and enhance prediction accuracy, suggesting that the true information dynamics are curved but not linear. Future quarterly abnormal earnings are generally statistically significant for estimating the valuation model. Our maximum likelihood forecast errors based on quarterly earnings are comparable to those documented in prior empirical research based on annual earnings, promising fruitful applications of quarterly forecasts for price prediction. Quarterly data exhibit mean reversion and stationarity, which are desirable attributes for estimation and prediction.

More importantly, our analyses demonstrate that, consistent with our expectation indeed, Bayesian forecasts are better than maximum likelihood forecasts. With transformations or without, our computed forecast errors of industrial firms under maximum likelihood average between 44.34% to 45.84%. These results are comparable to prior results. On the contrary, Bayesian errors are significantly smaller, averaging 18.20% without transformations, 12.22% with log transformation, and 11.99% with cubic root transformation. For financial firms, our Bayesian forecasts also have lower errors compared to the maximum likelihood results. Maximum likelihood errors average between 16.75% and 19.91%, while Bayesian errors average 9.29% without transformation, 11.62% with log transformation, and 9.60% with cubic root transformation.
Overall, we contribute to the literature by proposing methods to improve the implementation of the RIM, specifically by using transformations to explore non-linear relationships between price and accounting data, and by using Bayesian statistics to improve the inference mechanics. We show that the proposed methods yield forecasts that are significantly more accurate than those based on current implementations of the RIM.
Appendix A – The Residual Income Model

In economics and finance, the traditional approach to the problem of stock valuation based on a single firm has focused on the Dividend Discount Model (DDM) of Rubinstein (1976). It defines the value of a firm as the present value of the expected future dividends. That is, under the assumption of no arbitrage, there exists a pricing kernel \( \pi_{t+i} = (1 + r_t)^{-i} \) such that the price of a stock \( P_t \) is related to its dividends \( d_t \) by

\[
P_t = E_t \left[ \sum_{i=1}^{\infty} \pi_{t+i} d_{t+i} \right] = E_t \left[ \sum_{i=1}^{\infty} (1 + r_t)^{-i} d_{t+i} \right]
\]

(A1)

where

\( r_t \) denotes the discount rate during time period \( t \).

\( E_t[.] \) denotes the expectations operator conditioned on the date \( t \) information.

The idea of DDM implies that one should forecast dividends in order to estimate the stock prices. Since dividends are arbitrarily decided by management, it may be hard to estimate a dividend process in small samples. Moreover, market participants tend to focus on accounting information, especially earnings. The fundamental relation between book value of equity \( bv_{t} \), earnings \( x_t \) and dividends \( d_t \) is described in the Clean Surplus Accounting Model by equation \( bv_{t} = bv_{t-1} + x_t - d_t \), i.e., the change in book value equals earnings minus dividends. This relations is called the Clean Surplus Relation which can also be written as

\[
d_t = x_t - (bv_{t} - bv_{t-1})
\]

(A2)

Substituting equation (A2) to equation (A1), thereby eliminating dividends, yields the Residual Income Valuation Model (RIM) described in Peasnell (1982). It is a function of only accounting variables, namely:

\[
P_t = bv_t + \sum_{i=1}^{\infty} (1 + r_t)^{-i} (x_{t+i} - r_t bv_{t+i-1})
\]

(A3)

This principle shows that the theoretical value of the firm is equal to the opening book value of equity plus the present value of its residual income (or abnormal earnings, or excess earning), and will not be affected by accounting choices.

Ohlson (1995) states that the value of the firm can be well approximated even over a finite horizon by a function of forecasted earnings, book value, and discount rates. The only assumption required is that these forecasts be consistent with clean surplus relation. A variable \( V_t^T \) is defined as follows:

\[
V_t^T = bv_t + \frac{(1 + r_t)^T}{(1 + r_t)^T - 1} \sum_{i=1}^{T} (1 + r_t)^{-i} E_t[x_{t+i} - r_t bv_{t+i-1}]
\]

(A4)

Note that the amount \( V_t^T \) is a function of future earnings and book values measured over a finite horizon. However, despite the limited horizon, \( V_t^T \) approximates the value of the firm, so long as the horizon is “long enough”, which can be described as the following equation.
Equations (A4) and (A5) imply that the ability to predict earnings and book value --- even over a finite horizon --- is tantamount to the ability to approximate current value. If \( V^T_t \) provides a good approximation of firm value, we should be able to explain a large fraction of the variation in stock prices with the variables on the right-hand side of equation (A4). This suggests the following regression model:

\[
P_{it} = \beta_{i0} + \beta_{i1} bv_{at} + \sum_{t=1}^{T} \beta_{i,t+\tau} E_t [x_{i,t+\tau} - r_t bv_{i,t+\tau-1}] + v_{it} \tag{A7}
\]

We rewrite (A7) as:

\[
y_{it} = \beta_{i0} + \beta_{i1} bv_{it} + \sum_{k=1}^{n} \beta_{i,k+1} x_{i,t+k-1}^a + v_t = x_{i,t}^\prime \beta + v_t \tag{A8}
\]

\( k = 1, 2, 3, 4, \ldots, n; t = 1, \ldots, T. \)

Where \( x_{i,t+k-1}^a \) in (A8) is abnormal earnings, or \( E_t [x_{i,t+\tau} - r_t bv_{i,t+\tau-1}] \) in A(7).

In regression equation (A8), the AR(1) linear information dynamic, which assumes \( v_{it} \) follows an AR(1) process, can be written as

\[
v_{it} = \rho v_{i,t+1} + \varepsilon_{it} \tag{A9}
\]

Note that Ohlson’s linear information dynamic assumption is that abnormal earnings satisfy the following processes:

\[
x_{i,t+1}^a = \omega x_{i,t}^a + v_{it} + \varepsilon_{it+1}
\]

\[
v_{it} = \rho v_{i,t+1} + \varepsilon_{it}
\]

We depart from the Ohlson’s assumption in that we only consider the process described in (A9), to be directly concerned with valuation.
Appendix B – Procedures for Bayesian analysis
The procedures for Bayesian analysis are set up in five steps, setting the likelihood function, setting the posterior distribution of the parameters, Gibbs Sampling, assessing the model fit, and forecasting.

Step 1 Likelihood function
We first describe the observation
\[ y = (y_{11}, y_{12}, \ldots, y_{1T}, y_{21}, y_{22}, \ldots, y_{2T}, \ldots, y_{N1}, \ldots, y_{NT}) \]
by the parameters \( \{\beta, \mu, \rho, \sigma^2\} \), where \( N \) is the number of companies in a group, and \( T \) is the number of periods. Under the assumption that the observations are independent among the periods and among the companies, we can get the following likelihood function:
\[
p(y \mid \beta, \mu, \rho, \sigma^2) = \prod_{k=1}^{N} \prod_{t=2}^{T} N(y_{kt} \mid \beta + \mu, \rho + \sigma^2). \tag{B1}
\]

Step 2. Posterior distribution of the parameters
Second, we assign a prior distribution to each parameter.
\[
\begin{align*}
\pi(\beta) &= N_K(\theta, \Delta), \\
\pi(\mu) &= N(\mu_0, \sigma_0^2), \\
\pi(\rho) &= U(-1,1) \\
\pi(\sigma^2) &= Inv - gamma(a, b). \tag{B2}
\end{align*}
\]
where

1. \( \theta = B = (X'X)^{-1}X'y \), where \( X = \left( x_{-11}, x_{-12}, \ldots, x_{-1T}, \ldots, x_{-N1}, \ldots, x_{-NT} \right) \).
2. \( \Delta = 100(X'X)^{-1} \frac{SS_E}{NT - p} \), where \( SS_E = y'y - 2B'X'y + B'X'XB \), \( N \) is the number of companies, \( T \) is the number of observations and \( p \) is the degree of freedom.
3. \( \mu_0 = (NT)^{-1} \sum_{j=1}^{N} \sum_{t=1}^{T} (y_{ij} - x_{ij} B) \).
4. \( \sigma_0^2 = \frac{SS_E}{NT - p} \)
5. \( a = b = 0.001 \).

The idea in choosing those hyperparameters \( (\theta, \Delta, \mu_0, \sigma_0^2) \) is to use the estimation of parameters from the ordinary linear regression model
\[ y_{it} = x_{it}' \beta + \varepsilon_{it}, \varepsilon_{it} \sim iid \ N(0, \sigma^2), i = 1, \ldots, N; t = 1, \ldots, T. \]
Assume that all the parameters are independent of each other, the joint distribution of the parameters can be expressed as
\[
p(\beta, \mu, \rho, \sigma^2) = N_p(\beta | \theta, \Delta) \cdot N(\mu | \mu_0, \sigma^2_0) \cdot U(\rho | -1,1) \cdot \Gamma(\sigma^2 | a,b), \tag{B3}
\]
where P is the number of predictors.

From the likelihood function in (B1) and joint prior distribution in (B3), we derive the posterior distribution of the parameters by Bayes’ rule:
\[
p(\beta, \mu, \rho, \sigma^2 | y) \propto \prod_{i=1}^{N} \mathcal{N}(y_{it} | \hat{x}_i \beta + \mu, \sigma^2) \cdot \prod_{i=2}^{T} \mathcal{N}(y_{it} | \hat{x}_i \beta + \rho \left( y_{i,t-1} - \hat{x}_{i,t-1} \beta \right), \sigma^2) \tag{B4}
\]
\[
\cdot N_p(\beta | \theta, \Delta)N(\mu | \mu_0, \sigma^2_0)U(\rho | -1,1)\Gamma(\sigma^2 | a,b)
\]

**Step 3. Gibbs Sampling**

The Gibbs sampler is an iterative Monte Carlo algorithm designed to extract the posterior distribution from the tractable complete conditional distributions rather than directly from the intractable joint posterior distribution, which is difficult to acquire in explicit form. The target is to make inferences on the parameters \( \{\beta, \mu, \rho, \sigma^2\} \) given the data. We consider the complete conditional distributions \( \pi(\beta | .), \pi(\mu | .), \pi(\rho | .), \) and \( \pi(\sigma^2 | .) \) respectively. Here, the conditioning argument “.” denotes the observation \( y \) and the remaining parameters.

The Gibbs sampler is employed by the following six steps:

1. Step 1, obtain starting values \( (\beta^0, \mu^0, \rho^0, \sigma^{2,0}) \).
2. Step 2, draw \( \beta^t \) from \( \pi(\beta | \mu^{t-1}, \rho^{t-1}, \sigma^{2,t-1}, y) \).
3. Step 3, draw \( \mu^t \) from \( \pi(\mu | \beta^t, \rho^{t-1}, \sigma^{2,t-1}, y) \).
4. Step 4, draw \( \rho^t \) from \( \pi(\rho | \beta^t, \mu^t, \sigma^{2,t-1}, y) \).
5. Step 5, draw \( \sigma^{2,t} \) from \( \pi(\sigma^2 | \beta^t, \mu^t, \rho^t, y) \).
6. Step 6, repeat a large number of times. (In this paper, 11000 iterations are run in the Gibbs sampling, the first 1000 draws are thrown away, and 1000 draws are collected by picking one draw every 10 paces. These 1000 draws are then trimmed at the 90th and 10th percentiles.)

The complete conditional distributions for the parameters are
\[
(\alpha) \beta | y, \mu, \rho, \sigma^2 \sim N_p(\beta | (I - \Lambda)\theta + \Lambda \mu, \Lambda \Sigma \beta), \text{ where}
\]
\[
\mu - \beta = \left[ \sum_{i=1}^{N} \left( x_{-i}^i \beta_i + \sum_{t=2}^{T} \left( x_{-i}^t \rho x_{-t,i,t-1} \beta_i - \rho x_{-t,k,t-1} \right) \right) \right]^{-1} \\
\cdot \left[ \sum_{i=1}^{N} \left( y_{-i} - \mu - \sum_{t=2}^{T} 2(y_{-i} - \rho y_{i,t-1}) \beta_i - \rho y_{i,t-1} \right) \right],
\]

\[
\Sigma_{\beta} = \sum_{i=1}^{N} \left( x_{-i}^i \beta_i \left( \sum_{i=1}^{N} \left( x_{-i}^i \beta_i \right) - \rho x_{-i} \beta_i \right) \right),
\]

\[
\Lambda = (\Delta^{-1} + \Sigma_{\beta}^{-1})^{-1} = (\Sigma_{\beta}^{-1} + \Sigma_{\beta}^{-1})^{-1} = (\Sigma_{\beta} + \Delta)^{-1}.
\]

(b) \[\mu \mid y, \beta, \rho, \sigma^2 \sim N \left( \mu \left( 1 - \Phi \right) \mu_0 + \Phi \sum_{i=1}^{N} \left( y_{-i} - \rho x_{-i} \beta \right), \Phi \sigma^2 \right),\] where \[\Phi = -\frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}.
\]

(c)
\[
\rho \mid y, \beta, \mu, \sigma^2 \sim U(\rho \mid -1,1)N \left( \sum_{i=1}^{N} \sum_{t=2}^{T} \left( y_{-i}^t - x_{-i}^t \beta \right) \left( y_{i,t-1} - x_{-i,t-1} \beta \right), \sum_{i=1}^{N} \sum_{t=2}^{T} \left( y_{i,t-1} - x_{-i,t-1} \beta \right)^2 \right)
\]

(d) \[\sigma^2 \mid y, \beta, \mu, \rho \sim \Gamma \left( a + \frac{TN}{2}, b + \frac{S}{2} \right),\]
where
\[
S = \sum_{i=1}^{N} \left( y_{-i} - \rho x_{-i} \beta - \mu \right)^2 + \sum_{i=1}^{N} \sum_{t=2}^{T} \left( y_{i,t} - x_{-i,t} \beta - \rho y_{i,t-1} - x_{-i,t-1} \beta \right)^2.
\]

We choose the starting points \((\beta^0, \rho^0, \mu^0, \sigma^{2,0})\) as follows:

(6) \[\beta^0 = (X'X)^{-1} X'y.\]

(7) \[\rho^0 = \frac{SS_{12}}{\sqrt{SS_{11}SS_{22}}}, \] \[SS_{12} = \sum_{i=1}^{N} \sum_{t=2}^{T-1} (y_{-i} - x_{-i} \beta^0 - ave) (y_{i,t} - x_{-i,t} \beta^0 - ave2),\]

\[SS_{11} = \sum_{i=1}^{N} \sum_{t=1}^{T-1} (y_{-i} - x_{-i} \beta^0 - ave)^2,\]

\[SS_{22} = \sum_{i=1}^{N} \sum_{t=2}^{T} (y_{i,t} - x_{-i,t} \beta^0 - ave2)^2,\]

\[avel = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T-1} (y_{-i} - x_{-i} \beta^0)}{N(T-1)}, \text{ave2} = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} (y_{i,t} - x_{-i,t} \beta^0)}{N(T-1)}.\]

(8) \[\mu^0 = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{i,t} - x_{-i,t} \beta^0).\]

(9) \[\sigma^{2,0} = \frac{SS_E}{NT - p}, \text{where} \ SS_E = y'y - 2 (\beta^0)' X'y + (\beta^0)' X'X \beta^0.\]
The ideas of setting $\beta^0, \mu^0$ and $\sigma^{2.0}$ are from the estimation of parameters from the ordinary linear regression model $y_{it} = x_{it}'\beta + \epsilon_{it}, \epsilon_{it} \sim N(0, \sigma^2), i = 1, \ldots, N; t = 1, \ldots, T$. The idea of setting $\rho^0$ is using it as the autocorrelation of time series $\nu_{it,i,t-1} + \epsilon_{it}$ in an AR(1) structure.

**Step 4. Estimating Model Fit**

In this chapter, the conditional predictive ordinate is defined as

$$p(y_{i,t+1} | y_{(it)}) \approx \sum_{h=1}^{M} \omega_{it}^{(h)} p(y_{i,t+1} | \Omega^{(h)}, y_{(it)})$$

where $y_{i,t+1}$ denotes the random future observation of company i at period t + 1, $y_{(it)} = (y_{i1}, y_{i2}, \ldots, y_{it})$ denotes the observations of company i from period 1 to t, $\Omega^{(h)}$ denotes the hth draw of the parameters from the Gibbs sampler, and

$$\omega_{it}^{(h)} = \frac{f(y_{(it)} | \Omega^{(h)})}{\sum_{k=1}^{M} f(y_{(it)} | \Omega^{(k)})}$$

$k = 1, \ldots, M$.

**Step 5. Forecast**

After getting the posterior distribution of the parameters, we can use it to predict the future stock prices at period $T + 1$ for each firm in the group $y_{i,T+1}, i = 1, 2, \ldots, N$, given the data $y_{(iT)} = (y_{i1}, y_{i2}, \ldots, y_{iT}), i = 1, 2, \ldots, N$. The predictions can be sampled from the posterior predictive distribution

$$f(y_{i,T+1} | y_{(iT)}) = \int f(y_{i,T+1} | \Omega) \pi(\Omega | y) d\Omega$$

where $\Omega = (\mu, \beta, \rho, \sigma^2)$.

Let $\Omega^{(1)}, \Omega^{(2)}, \ldots, \Omega^{(M)}$ be a sequence of range M from the Gibbs sampler, then an estimation of $f(y_{i,T+1} | y_{(iT)})$ is

$$\hat{f}(y_{i,T+1} | y_{(iT)}) = M^{-1} \sum_{h=1}^{M} f(y_{i,T+1} | \Omega^{(h)}, y_{(iT)}).$$

(B7)
To get samples of $y_{i,T+1}$, we use data argumentation to fill in $y_{i,T+1}$ to each $\Omega^{(h)}$, $h = 1, 2, \cdots, M$ to get $y_{i,T+1}^{(h)}$, $h = 1, 2, \cdots, M$ from the normal distribution described below. 

$$y_{i,T+1} | \Omega, y_{i,T} \sim N(\hat{x}^{'}_{i,T+1} \beta + \rho(y_{i,T} - \hat{x}^{'}_{i,T} \beta), \sigma^2) \quad (B8)$$
<table>
<thead>
<tr>
<th>Model</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>$y_t = \beta_0 + \beta_1 b v_t + \sum_{k=1}^{4} \beta_{k+1} x_{t+k}^a + v_t, v_t = \rho v_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$</td>
</tr>
<tr>
<td>AR(2)</td>
<td>$y_t = \beta_0 + \beta_1 b v_t + \sum_{k=1}^{4} \beta_{k+1} x_{t+k}^a + v_t, v_t = \rho_1 v_{t-1} + \rho_2 v_{t-2} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$</td>
</tr>
<tr>
<td>Log - AR(1)</td>
<td>$\log(y_t) = \beta_0 + \beta_1 b v_t + \sum_{k=1}^{4} \beta_{k+1} x_{t+k}^a + v_t, v_t = \rho v_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$</td>
</tr>
<tr>
<td>Square Root - AR(1)</td>
<td>$\sqrt{y_t} = \beta_0 + \beta_1 b v_t + \sum_{k=1}^{4} \beta_{k+1} x_{t+k}^a + v_t, v_t = \rho v_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$</td>
</tr>
<tr>
<td>Cubic root - AR(1)</td>
<td>$\sqrt[3]{y_t} = \beta_0 + \beta_1 b v_t + \sum_{k=1}^{4} \beta_{k+1} x_{t+k}^a + v_t, v_t = \rho v_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$</td>
</tr>
<tr>
<td>Inverse - AR(1)</td>
<td>$1/y_t = \beta_0 + \beta_1 b v_t + \sum_{k=1}^{4} \beta_{k+1} x_{t+k}^a + v_t, v_t = \rho v_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$</td>
</tr>
<tr>
<td>Log - AR(2)</td>
<td>$\log(y_t) = \beta_0 + \beta_1 b v_t + \sum_{k=1}^{4} \beta_{k+1} x_{t+k}^a + v_t, v_t = \rho_1 v_{t-1} + \rho_2 v_{t-2} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$</td>
</tr>
<tr>
<td>Square Root - AR(2)</td>
<td>$\sqrt{y_t} = \beta_0 + \beta_1 b v_t + \sum_{k=1}^{4} \beta_{k+1} x_{t+k}^a + v_t, v_t = \rho v_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$</td>
</tr>
<tr>
<td>Cubic root - AR(2)</td>
<td>$\sqrt[3]{y_t} = \beta_0 + \beta_1 b v_t + \sum_{k=1}^{4} \beta_{k+1} x_{t+k}^a + v_t, v_t = \rho v_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$</td>
</tr>
<tr>
<td>Inverse - AR(2)</td>
<td>$1/y_t = \beta_0 + \beta_1 b v_t + \sum_{k=1}^{4} \beta_{k+1} x_{t+k}^a + v_t, v_t = \rho v_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$</td>
</tr>
</tbody>
</table>
Table 2: Descriptive Statistics of Raw Data

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>5%</th>
<th>25%</th>
<th>Median</th>
<th>75%</th>
<th>95%</th>
<th>Max</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price per share</td>
<td>1.60</td>
<td>8.92</td>
<td>21.78</td>
<td>33.01</td>
<td>46.31</td>
<td>71.50</td>
<td>293.56</td>
<td>35.87</td>
</tr>
<tr>
<td>Book Value per Share (BPS0)</td>
<td>0.09</td>
<td>1.86</td>
<td>4.84</td>
<td>8.66</td>
<td>14.20</td>
<td>28.25</td>
<td>59.24</td>
<td>10.84</td>
</tr>
<tr>
<td>Quarterly earnings forecast of</td>
<td>-1.33</td>
<td>0.00</td>
<td>0.18</td>
<td>0.36</td>
<td>0.58</td>
<td>1.09</td>
<td>3.15</td>
<td>0.43</td>
</tr>
<tr>
<td>the current quarter (EPS1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quarterly earnings forecast of</td>
<td>-0.42</td>
<td>0.02</td>
<td>0.20</td>
<td>0.38</td>
<td>0.61</td>
<td>1.12</td>
<td>2.63</td>
<td>0.45</td>
</tr>
<tr>
<td>one quarter ahead (EPS2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quarterly earnings forecast of</td>
<td>-0.36</td>
<td>0.03</td>
<td>0.22</td>
<td>0.40</td>
<td>0.63</td>
<td>1.15</td>
<td>2.60</td>
<td>0.47</td>
</tr>
<tr>
<td>two quarters ahead (EPS3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quarterly earnings forecast of</td>
<td>-0.46</td>
<td>0.05</td>
<td>0.23</td>
<td>0.41</td>
<td>0.65</td>
<td>1.18</td>
<td>2.56</td>
<td>0.49</td>
</tr>
<tr>
<td>three quarters ahead (EPS4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tbill rate of the current</td>
<td>0.92</td>
<td>0.92</td>
<td>1.25</td>
<td>1.96</td>
<td>4.74</td>
<td>6.02</td>
<td>6.02</td>
<td>2.97</td>
</tr>
<tr>
<td>quarter (R1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This table contains one observation for each of the 5328 firm-quarters in the total sample.
Table 3: Maximum Likelihood Estimation

<table>
<thead>
<tr>
<th>Panel A</th>
<th>[y_t = \beta_0 + \beta_1 b_{t-1} + \sum_{k=1}^{4} \beta_k v_{t-k} + \epsilon_t, \epsilon_t \sim N(0, \sigma^2)]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Industrial (N=172)</td>
</tr>
<tr>
<td>(\beta_0)</td>
<td>21.51 (p&lt;.0001)</td>
</tr>
<tr>
<td>(\beta_1)</td>
<td>-0.03 (p&lt;.6196)</td>
</tr>
<tr>
<td>(\beta_2)</td>
<td>8.48 (p&lt;.0001)</td>
</tr>
<tr>
<td>(\beta_3)</td>
<td>6.80 (p&lt;.0001)</td>
</tr>
<tr>
<td>(\beta_4)</td>
<td>11.03 (p&lt;.0001)</td>
</tr>
<tr>
<td>(\beta_5)</td>
<td>8.27 (p&lt;.0001)</td>
</tr>
<tr>
<td>(\rho)</td>
<td>-0.82 (p&lt;.0001)</td>
</tr>
<tr>
<td>Total R-square</td>
<td>75.87%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B</th>
<th>[\log(y_t) = \beta_0 + \beta_1 b_{t-1} + \sum_{k=1}^{4} \beta_k v_{t-k} + \epsilon_t, \epsilon_t \sim N(0, \sigma^2)]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Industrial</td>
</tr>
<tr>
<td>(\beta_0)</td>
<td>2.94 (p&lt;.0001)</td>
</tr>
<tr>
<td>(\beta_1)</td>
<td>0.003 (p=.0609)</td>
</tr>
<tr>
<td>(\beta_2)</td>
<td>0.24 (p&lt;.0001)</td>
</tr>
<tr>
<td>(\beta_3)</td>
<td>0.21 (p&lt;.0001)</td>
</tr>
<tr>
<td>(\beta_4)</td>
<td>0.32 (p&lt;.0001)</td>
</tr>
<tr>
<td>(\beta_5)</td>
<td>0.24 (p&lt;.0001)</td>
</tr>
<tr>
<td>(\rho)</td>
<td>-0.86 (p&lt;.0001)</td>
</tr>
<tr>
<td>Total R-square</td>
<td>82.55%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C</th>
<th>[\sqrt{3} y_t = \beta_0 + \beta_1 b_{t-1} + \sum_{k=1}^{4} \beta_k v_{t-k} + \epsilon_t, \epsilon_t \sim N(0, \sigma^2)]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Industrial</td>
</tr>
<tr>
<td>(\beta_0)</td>
<td>2.71 (p&lt;.0001)</td>
</tr>
<tr>
<td>(\beta_1)</td>
<td>0.00 (p=.2283)</td>
</tr>
<tr>
<td>(\beta_2)</td>
<td>0.26 (p&lt;.0001)</td>
</tr>
<tr>
<td>(\beta_3)</td>
<td>0.21 (p&lt;.0001)</td>
</tr>
<tr>
<td>(\beta_4)</td>
<td>0.34 (p&lt;.0001)</td>
</tr>
<tr>
<td>(\beta_5)</td>
<td>0.25 (p&lt;.0001)</td>
</tr>
<tr>
<td>(\rho)</td>
<td>-0.85 (p&lt;.0001)</td>
</tr>
<tr>
<td>Total R-square</td>
<td>81.39%</td>
</tr>
</tbody>
</table>

\(y_t\) is the stock price per share at time \(t\), \(b_{t-1}\) is the book value per share at the beginning of time \(t\), \(x_t^a\) represents the abnormal earning at time \(t\), \(\beta = (\beta_0, \cdots, \beta_n)^\top\) is the vector of intercept and slope coefficients of the predictors, and \(v_t\) is the other information term.
Table 4: Bayesian Estimation

<table>
<thead>
<tr>
<th>Panel A</th>
<th>Industrial</th>
<th>Financial</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_i = x_i' \beta + \mu + \epsilon_i,$</td>
<td>-4.47</td>
<td>-4.65</td>
</tr>
<tr>
<td>$\epsilon_i \sim N(0, \sigma^2),$</td>
<td>$t = 1, \ldots, T,$</td>
<td>$\pi(\beta) = N(\beta</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\pi(\mu) = N(\mu</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\pi(\rho) = U(\rho</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\pi(\sigma^2) = IG(\sigma^2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Log(CPO)</th>
<th>Industrial</th>
<th>Financial</th>
</tr>
</thead>
<tbody>
<tr>
<td>-4.26</td>
<td>-4.72</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B</th>
<th>Industrial</th>
<th>Financial</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_i = x_i' \beta + \rho(y_{i-1} - x_i' \beta) + \epsilon_i,$</td>
<td>-4.26</td>
<td>-4.72</td>
</tr>
<tr>
<td>$\epsilon_i \sim N(0, \sigma^2),$</td>
<td>$t = 1, \ldots, T,$</td>
<td>$\pi(\beta) = N(\beta</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\pi(\mu) = N(\mu</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\pi(\rho) = U(\rho</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\pi(\sigma^2) = IG(\sigma^2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Log(CPO)</th>
<th>Industrial</th>
<th>Financial</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3.10</td>
<td>-3.44</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C</th>
<th>Industrial</th>
<th>Financial</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_i = x_i' \beta + \rho(y_{i-1} - x_i' \beta) + \epsilon_i,$</td>
<td>-3.10</td>
<td>-3.44</td>
</tr>
<tr>
<td>$\epsilon_i \sim N(0, \sigma^2),$</td>
<td>$t = 1, \ldots, T,$</td>
<td>$\pi(\beta) = N(\beta</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\pi(\mu) = N(\mu</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\pi(\rho) = U(\rho</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\pi(\sigma^2) = IG(\sigma^2</td>
</tr>
</tbody>
</table>

$y_t$ is the stock price per share at time $t$, $b v_t$ is the book value per share at the beginning of time $t$, $x_t^a$ represents the abnormal earning at time $t$, $\beta = (\beta_0, \cdots, \beta_n)'$ is the vector of intercept and slope coefficients of the predictors, $x_t' = (1, b v_t, x_{t+1}^a, x_{t+1}^a, x_{t+3}^a, x_{t+4}^a, \cdots)'$ is the vector of intercept and predictors, and $v_t$ is the other information term. Log(CPO) is the log of conditional predictive ordinate. A larger Log(CPO) denotes better fit.
Table 5 – Absolute Prediction Error

<table>
<thead>
<tr>
<th></th>
<th>Un-transformed</th>
<th>Log</th>
<th>Cubic Root</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Industrial (N=172)</td>
<td>Financial (N=44)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
<td>Mean</td>
</tr>
<tr>
<td>MLE</td>
<td>45.84%</td>
<td>20.91%</td>
<td>16.75%</td>
</tr>
<tr>
<td>Bayesian</td>
<td>18.20%</td>
<td>10.27%</td>
<td>9.29%</td>
</tr>
<tr>
<td>P-value</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0050</td>
</tr>
</tbody>
</table>

Absolute prediction error is $R = |\frac{\hat{y}_{24} - y_{24}}{y_{24}}|$, where $\hat{y}_{24}$ and $y_{24}$ are respectively the predicted and actual price per share of a company for the 24th quarter. $R$ close to zero means low prediction error, or high accuracy.
References:


