Latent Bayesian Inference for Robust Earnings Estimates

Chirag Nagpal[*,†], **Robert E. Tillman**[*], **Prashant P. Reddy**[*], **Manuela M. Veloso**[*]
[*]J.P. Morgan AI Research
[†]Carnegie Mellon University
chiragn@cs.cmu.edu
{robert.e.tillman, prashant.reddy, manuela.veloso}@jpmorgan.com

Abstract

Equity research analysts at financial institutions play a pivotal role in capital markets; they provide an efficient conduit between investors and companies' management and facilitate the efficient flow of information from companies, promoting functional and liquid markets. However, previous research in the academic finance and behavioral economics communities has found that analysts' estimates of future company earnings and other financial quantities can be affected by a number of behavioral, incentive-based and discriminatory biases and systematic errors, which can detrimentally affect both investors and public companies. We propose a Bayesian latent variable model for analysts' systematic errors and biases which we use to generate a robust bias-adjusted consensus estimate of company earnings. Experiments using historical earnings estimates data show that our model is more accurate than the consensus average of estimates and other related approaches.

1 Introduction

Equity research analysts at financial institutions play a pivotal role in capital markets. Both large institutions and small investors alike lack the time and resources to analyze thousands of companies and meet with their management. Such activity would similarly overwhelm even the largest public companies. Analysts provide a necessary conduit between the companies' management and investors. In addition, regulators have a strong interest in the efficient flow of information from companies to ensure functional, liquid markets; analysts significantly contribute to this process [Bradshaw, 2011].

One important function of analysts is forecasting companies' future earnings and other financial quantities, e.g. revenue and cash flow. Investors use these quantities to understand the financial health of companies and the value of their stock. There are hundreds of academic studies which have analyzed these forecasts and their accuracy. Analysts' estimates of upcoming earning have been shown to be more accurate than those produced by a range of time-series models of actual reported earnings [Brown and Rozeff, 1978, Fried and Givoly, 1982]. Brown et al. [1987] finds that this is due to both a timing and information advantage. Accordingly, there is a strong relationship between analysts' revisions to their estimates and changes in stock prices [Michaely and Womack, 2005].

Despite the value in these estimates, the academic finance and behavioral economics communities have found significant empirical evidence of various types of analyst biases and systematic errors. Much of this literature focuses on over-optimism in analysts' earnings forecasts [De Bondt and Thaler, 1990, Hong and Kubik, 2003, Elliott et al., 2010]. Common explanations for this optimism include incentives for analysts to maintain good relations with companies' management [Francis and Philbrick, 1993, Richardson et al., 1999] and possible conflicts of interest with companies who are also investment banking clients [Michaely and Womack, 1999, O' Brien et al., 2005]. Other literature, however, has highlighted that analysts are actually pessimistic or conservative in their estimates

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

at shorter horizons and overly optimistic at longer-term horizons [Richardson et al., 1999, Brown, 1997, Eames and Glover, 2003, Elliott et al., 2010]. This literature explains this asymmetry similarly in terms analysts incentives to maintain good relations with companies' management: pessimistic projections for an upcoming quarter will result in a company beating expectations and a subsequent positive stock price reaction, whereas unrealistically optimistic projections will ultimately result in a company missing expectations and a negative stock price reaction.

Many papers have investigated more specific behavioral biases [De Bondt and Thaler, 1990] and asymmetries [Easterwood and Nutt, 1999] in analysts' earnings estimates. Elliott et al. [2010] finds that analysts systematically underweight new information and underreact to their own previous revisions. Easterwood and Nutt [1999] finds that analysts underreact to negative information and overreact to positive information. Ciccone [2005] shows that forecast errors are different for profit-making companies than loss-making companies. Raedy et al. [2006] provides evidence that analysts are less likely to make revisions in the opposite direction of previous revisions as they incur a greater reputational cost for acknowledging they "overshot" in previous estimates. There is also evidence analysts engage in "herding behavior" [O'Neill et al., 2011], or are unlikely to deviate too much from other analysts. Hong et al. [2000] finds that analysts are more likely to engage in herding in their revisions when their outstanding estimates deviate significantly from other analysts.

Furthermore, there is evidence that biases and systematic errors are stronger in some analysts than others. Michaely and Womack [1999] finds that analysts with more experience have lower forecast error. Hong et al. [2000] finds that experienced analysts are also more likely to revise their estimates earlier and their estimates show more dispersion, indicating stronger conviction and willingness to issue bold forecasts that go against the trend. Mikhail et al. [1999] finds that earnings estimate accuracy is negatively correlated with analyst turnover. Finally, recent research has highlighted evidence of racial, gender and political biases towards out of group company CEOs [Jannati et al., 2019].

Many investors use the *consensus earnings estimate*, or the simple average of estimates from analysts at major institutions, as a reliable forecast of a company's upcoming earnings. However, research has shown that inversely weighting these individual analysts' estimates based on their historical forecasting errors results in a more accurate "adjusted" consensus estimate [Jha and Mozes, 2001, Michaely et al., 2018]. While this is consistent with the finding that some analysts are more accurate than others, only considering the total forecasting error does not disentangle the systematic bias component of forecasting error from the unsystematic (or unpredictable) general error and does account for different types of biases and asymmetries in forecast errors. We hypothesize that a model which accounts for the systematic and asymmetric components of analyst estimate error will produce more robust adjusted consensus earnings estimates.

1.1 Primary contribution

We propose a Bayesian latent variable model for analysts' earnings estimate forecasting error and show how it can be used to infer a robust adjusted consensus earnings estimate. In our model, we assume there are latent subgroups of analysts such that analysts within each group demonstrate similar systematic forecasting errors. We use historical analyst estimates of company earnings and actual reported earnings to learn the parameters of this model. We then describe a procedure for inverse inference to generate a robust consensus estimate of future earnings from individual analysts' estimates. We believe robust earnings estimates benefit both investors, who require accurate forecasts of company financials, and public companies, whose stock prices may be undervalued as a result of some analysts' incentive biases, conflicts of interest or for discriminatory reasons.

In the following sections, we describe the proposed Bayesian latent variable model and inverse inference procedure to generate robust company earnings estimates. We compare the resulting robust estimates to actual reported company earnings. We find that this approach produces estimates which are more accurate than the consensus estimate and other adjusted consensus baselines.

2 Proposed Model for Robust Earnings Estimates

The specific financial quantity we focus on modeling the forecast error of is *Earnings per Share (EPS)*, as this most widely considered quantity when assessing the value of a public company. EPS is the ratio between company's net income during a particular reporting period after subtracting preferred dividends and the number of outstanding common shares of that company's stock.

We assume that there exists a latent subgrouping of analysts such that within each group, we observe similar EPS forecasting errors. We hypothesize that analysts' estimates for *changes* in EPS (the difference between the EPS forecast for the next period and the company's reported EPS from the previous period) are normally distributed around a linear function of the actual resulting change in EPS with some heteroscedastic variance. Both the variance and the parameters of the linear function are conditioned on the latent subgroup an analyst belongs to. In this setting, the distribution of the *forwards model*, or the observed forecasting error process, can be written in closed form and parameter learning can be carried out with a gradient-based method.

2.1 The Forwards Model

For a particular reporting period, we use X_i to represent the actual change in reported EPS from the previous period, where i is an index over the set of companies, $\mathcal{S}$, and ξ_i is an indicator of whether the change in EPS is positive or negative.

Now, for each analyst indexed by j in the set of analysts, $\mathcal{A}$, we draw a categorical variable z_j conditioned on the parameters θ_j that determines which one of the K latent subgroups analyst j belongs to. Finally we draw a set of parameters w_j conditioned on the analyst subgroup that interact with the true change in EPS X_i and ξ_i. We show the model in plate notation in Figure 1 and give the explicit steps in the process below:

1. For all $j \in \mathcal{A}$, Draw $z_j|\theta_j$ as,

$$z_j|\theta_j \sim \text{Discrete}(\textsc{Soft-Max}(\theta_j))$$

2. For all $i \in \mathcal{S}$, Draw $\xi_i|X_i$ as,

$$\xi_i = \mathbb{1}\{X_i > 0\}$$

3. For all $i \in \mathcal{S}$ and $j \in \mathcal{A}$,
Draw $\hat{X}_{ij}|X_i, \{z_j\}_{j=1}^{\mathcal{A}}\{\omega_k\}_{k=1}^{K}$ as

$$\hat{X}_{ij}|\cdot \sim \mathcal{N}(\alpha_{z_j} \cdot X_i + \beta_{zj}\ ,\ \sigma^2_{z_j}) \quad (1)$$

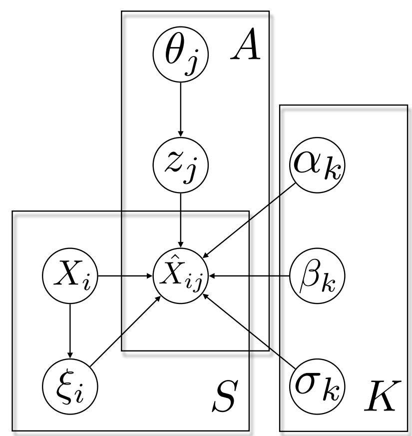

Figure 1: The model in plate notation. $\mathcal{A}$, $\mathcal{S}$, and $\mathcal{K}$, represent the analysts, companies, and latent subgroups. X_i is the actual change in EPS for company i between reporting periods and $\hat{X}_{ij}$ is analyst j's estimated change.

2.2 Parameter Learning

To learn the parameters Θ of the model from observed analyst estimates and actual reported EPS data $\mathcal{D}$, we want to maximize the likelihood, $P(\mathcal{D}|\Theta)$, which we can write as follows:

$$\begin{aligned} P(\mathcal{D}|\Theta) &= P(\{\{\hat{X}_{ij}\}_{j=1}^{|\mathcal{A}|}\}_{i=1}^{|\mathcal{S}|}, \{X_i, \xi_i\}_{i=1}^{|\mathcal{S}|} | \{\theta_j\}_{j=1}^{|\mathcal{A}|}, \{\alpha_k, \beta_k, \sigma_k\}_{k=1}^{K}) \\ &= \prod_{i=1}^{|\mathcal{S}|}\prod_{j=1}^{|\mathcal{A}|}\sum_{k=1}^{K} P_k(\hat{X}_{ij}|X_i, \xi_i, z_j)P(z_j|\theta_j) \qquad \text{(Here, } P_k \text{ is as in Eq. 1)} \\ \log P(\mathcal{D}|\Theta) &= \sum_{i=1}^{|\mathcal{S}|}\sum_{j=1}^{|\mathcal{A}|}\log\sum_{k=1}^{K} P_k(\hat{X}_{ij}|X_i, \xi_i, z_j)P(z_j|\theta_j) \end{aligned}$$

$$
\begin{aligned}
\log P(\mathcal{D}|\Theta) &= \sum_{i=1}^{|\mathcal{S}|}\sum_{j=1}^{|\mathcal{A}|} \log \mathbb{E}_{z_j|\theta_j}\left[P_k(\hat{X}_{ij}|X_i,\xi_i,z_j)\right] \\
&\geq \sum_{i=1}^{|\mathcal{S}|}\sum_{j=1}^{|\mathcal{A}|} \mathbb{E}_{z_j|\theta_j}\left[\log P_k(\hat{X}_{ij}|X_i,\xi_i,z_j)\right] \quad \text{(From Jensen's Inequality)} \\
&\triangleq \mathbf{ELBO}(\Theta;\mathcal{D})
\end{aligned}
$$

Here, $\mathbf{ELBO}(\Theta;\mathcal{D})$ is the *Evidence Lower Bound*, which is commonly used when carrying out Variational Inference with Graphical Models [Blei et al., 2017]. In the vernacular of Variational Inference, our approximating distribution, $q(Z)$ of the True Posterior is the posterior distribution of Z given θ, $p(Z_i|\theta_i)$.

2.3 Optimization

For parameter learning we utilize the popular First Order optimizer Adam [Kingma and Ba, 2015], which has a learning rate of 1×10^{-4}. We do not perform any minibatching and stop optimization as soon as we overfit the validation set.

The **ELBO** as defined in the previous section is explicitly optimized in the following form:

$$
\mathbf{ELBO}(\theta;\mathcal{D}) = -\sum_{i=1}^{|\mathcal{S}|}\sum_{j=1}^{|\mathcal{A}|}\sum_{k=1}^{K} \textsc{Soft-Max}^{(k)}(\theta_j)\cdot\left(\frac{||(\alpha_k^{(\xi_i)}\cdot\hat{X}_{ij}+\beta_k^{(\xi_i)})-X_i||_2^2}{\sigma_k}\right)
$$

2.4 Identification

For purposes of identification and to ensure convergence to a good local minimum, we fix the hyper parameters corresponding to the latent group $K = 1$ as $\alpha = 1$ and $\beta = 0$. Thus, the semantic interpretation of this latent group is that estimates from this group are accurate and unbiased.

2.5 Parameter Initialization

For all subgroups $K \neq 1$, we initialize α_k and β_k using the coefficient of Ridge regression estimates by regressing the estimates $\{\{\hat{X}_{ij}\}_{j=1}^{|\mathcal{A}|}\}_{i=1}^{N}$ on the corresponding set of changes in actual EPS $\{X_i\}_{i=1}^{N}$. We further set the initial variance of each latent group σ_k^2 to 1.0. We observe that in practice using these initial values leads to better convergence.

3 Inverse Inference for Generating Robust Estimates

At test time, we want to infer a robust estimate for change in EPS from the analysts' estimates X_{ij} and the learned model parameters Θ. For a company i, this is equivalent of inferring $P(x_i|\{\hat{x}_{ij}\}_{j=1}^{|\mathcal{A}|},\Theta)$.

In our formulation, inference at test time is harder than parameter learning. This is primarily because the posterior over the latent variables is intractable. We can, however, express the conditional distributions of each variable in closed form, which allows us to use Gibbs Sampling, a Markov Chain Monte Carlo technique that allows inference by sampling from the conditional distributions, to overcome this challenge. Sampling from the full conditionals is easy for all of the variables except the changes in EPS actuals X_i. Proposition 1 gives the posterior distribution of X_i given the analyst estimates, $\{\hat{X}_{ij}\}_{j=1}^{|\mathcal{A}|}$, and model parameters, $\{\alpha_k,\beta_k,\sigma_k\}_{k=1}^{K}$, in closed form to allow sampling.

Proposition 1 *Under the DAG model assumptions in Figure 1, the Posterior Distribution of X_i conditioned on its Markov Blanket, or set of variable such that X_i is conditionally independent of all other variables in the model, is given as*

$$X_i|[\widetilde{\mathbf{X}}], \{z_j\}_{j=1}^{|\mathcal{A}|} \sim \text{Multivariate-Normal}\Big([\boldsymbol{\mu}]\ , [\boldsymbol{\Sigma}]\Big)$$

where,

$$[\boldsymbol{\Sigma}] = \big(\sigma_0 + [\boldsymbol{\alpha}]^\top[\boldsymbol{\alpha}]\big)^{-1}, [\boldsymbol{\mu}] = [\boldsymbol{\Sigma}]([\boldsymbol{\alpha}]^\top[\widetilde{\mathbf{X}}])$$

and,

$$[\widetilde{\mathbf{X}}] = \begin{bmatrix} \hat{X}_{i0} - \beta_{z_0} \\ \hat{X}_{i1} - \beta_{z_1} \\ \vdots \\ \hat{X}_{i|\mathcal{A}|} - \beta_{z_{|\mathcal{A}|}} \end{bmatrix}, [\boldsymbol{\alpha}] = \begin{bmatrix} \alpha_{z_0} \\ \alpha_{z_1} \\ \vdots \\ \alpha_{z_{|\mathcal{A}|}} \end{bmatrix}, [\boldsymbol{\sigma}^2] = \begin{bmatrix} \sigma_{z_0}^2 & & & \\ & \sigma_{z_1}^2 & & \\ & & \ddots & \\ & & & \sigma_{z_{|\mathcal{A}|}}^2 \end{bmatrix}$$

Proof Sketch.

The Proof of the following proposition, involves adding a weak conjugate prior on $X_i \sim \mathcal{N}(0, \sigma_0)$. Now,

$$\hat{X}_{ij}|X_i, z_j \sim \mathcal{N}(\alpha_{z_j}{\cdot}X_i + \beta_{z_j}\ ,\ \sigma_{z_j}^2)$$

$$\hat{X}_{ij} - \beta_{z_j}|X_i, z_j \sim \mathcal{N}(\alpha_{z_j}{\cdot}X_i\ ,\ \sigma_{z_j}^2)$$

$$\{\hat{X}_{ij} - \beta_{z_j}\}_{j=1}^{|\mathcal{A}|}|X_i, \{z_j\}_{j=1}^{|\mathcal{A}|} \sim \prod_{j=1}^{|\mathcal{A}|} \mathcal{N}(\alpha_{z_j}{\cdot}\ X_i\ ,\ \sigma_{z_j}^2)$$

Rewriting in matrix form, we get

$$[\widetilde{\mathbf{X}}]|X_i, \{z_j\}_{j=1}^{|\mathcal{A}|} \sim \text{Multivariate-Normal}\left(X_i^\top[\boldsymbol{\alpha}]\ , [\boldsymbol{\sigma}^2]\right) \tag{2}$$

Now, from Equation 2 and the result in Cepeda and Gamerman [2000, 2005] pertaining to Bayesian Linear Regressions under Heteroscedasticity, we arrive at the posterior. ■

Algorithm 1 provides the steps in the Gibbs sampling procedure for inverse inference of X_i by sampling from the full posterior conditionals.

Algorithm 1: Gibb's Sampler for $P(\{X_i, \xi_i\}_{i=1}^{|\mathcal{S}|}|\cdot)$

Input: $\{\{\hat{X}_{ij}\}_{i=1}^{|\mathcal{S}|}\}_{j=1}^{|\mathcal{A}|}, \{\alpha_k, \beta_k, \sigma_k\}_{k=1}^{K}, \{\theta_j\}_{j=1}^{|\mathcal{A}|}$
Initialize:
for $i \leftarrow 1$ **to** $|\mathcal{S}|$ **do**
  $X_i^{(0)} = \frac{1}{|\mathcal{A}|}\sum_j \hat{X}_{ij}$; (Consensus Estimate)
  $\xi_i^{(0)} = \mathbb{1}\{X_i^{(0)} > 0\}$;
end
for $n \leftarrow 1$ **to** N **do**
  for $j \leftarrow 1$ **to** $\mathcal{A}$ **do**
    $z_j^{(n)} \sim$ Discrete(Soft-Max(θ_j));
  end
  for $i \leftarrow 1$ **to** $\mathcal{S}$ **do**
    $X_i^{(n)} \sim P_{\text{posterior}}(X_i^{(n)}|\{z_j^{(n)}\}_{j=1}^{|\mathcal{A}|}, \xi_i^{(n)}, \{\alpha_k, \beta_k, \sigma_k\}_{k=1}^{K})$; (As in Prop. 1)
    $\xi_i^{(n)} = \mathbb{1}\{X_i^{(n)} > 0\}$;
  end
end
Output: $\{X_i^{(0)}, X_i^{(1)}, \cdots, X_i^{(1)}\}$

We evaluate this procedure using historical EPS estimates and actuals in the next section.

4 Experiments

We evaluate the ability of our model and inverse inference procedure to generate robust consensus estimates by first learning parameters from historical EPS estimates and actuals and then carrying out inverse inference to predict changes in EPS actuals X_i from test data estimates X_{ij}. We compare these estimates using our approach, which we refer to as **Latent Bayesian Averaging (LBA)**, to the simple consensus estimates and other reference baselines, which we describe below.

4.1 Reference Baselines

We consider the following reference baselines to benchmark our approach:

No Adjustment (NA): The estimate of X_i is the simple consensus estimate, or average of all of the analysts' estimates $\{\hat{X}_{ij}\}_{j=1}^{|\mathcal{A}|}$.

$$\hat{X}_i = \frac{1}{|\mathcal{A}|}\sum_{j=1}^{|\mathcal{A}|}\hat{X}_{ij}$$

Weighted Adjustment (WA): Instead of averaging over the estimates, $\{\hat{X}_{ij}\}_{j=1}^{|\mathcal{A}|}$ naively, we perform a weighted averaging such that $w_j \propto \frac{1}{\text{error}_j}$, i.e. the weight given to an analyst's estimate is inversely proportional to the analyst's historical forecast accuracy.

$$\hat{X}_i = \frac{1}{|\mathcal{A}|}\sum_{j=1}^{|\mathcal{A}|} w_j \cdot \hat{X}_{ij}$$

Regression Adjustment (RA): We regress the set of true values, $\{X\}_{i=1}^{N}$ against the corresponding estimates, $\{\hat{X}\}_{a=1}^{|\mathcal{A}|}$. At test time, we perform the learnt regression on $\{\hat{X}\}_{a=1}^{|\mathcal{A}|}$ to get adjusted estimates for X. The final estimate is the average of the adjusted estimates.

We consider two different regression functions, a parametric ridge regression **RA-Ridge** and a non-parametric regression consisting of an Random Forest of Decision Trees **RA-Ensemble**.

$$\hat{X}_i = \frac{1}{|\mathcal{A}|}\sum_{j=1}^{|\mathcal{A}|} f(\hat{X}_{ij}, \theta)$$

where $f(.)$ is the learnt regression function.

Bayesian Regression Adjustment (BA) : Instead of regressing the actual X_i on the estimates, $\{\hat{X}_{ij}\}_{j=1}^{|\mathcal{A}|}$. We first learn a regression of $\{\hat{X}_{ij}\}_{j=1}^{|\mathcal{A}|}$ on the actuals X_i with a linear link function $f(\beta^\top X_i)$. At test time we condition on β and place a weak conjugate prior on X_i. The final adjusted estimate of X_i is then recovered as the expectation of X_i under the posterior conditioned on $\{\hat{X}_{ij}\}_{j=1}^{|\mathcal{A}|}$ and β.

$$\hat{X}_i = \mathbb{E}_{X_i \sim p(.|\{\hat{X}_{ij}\}_{j=1}^{|\mathcal{A}|}, \beta)}[X_i]$$

Note that **Bayesian Regression Adjustment** is equivalent to our model when the number of latent groups $K = 1$.

4.2 Dataset

We use the Thompson Reuters' **Institutional Brokers Estimate System (I/B/E/S)** Dataset which records estimated earnings forecasts of different analysts for different companies and upcoming periods across multiple time horizons. For our experiments we look at a smaller subset of the **I/B/E/S** data consisting of the top 200 companies followed by the most analysts over a 19 year period from

January 1st, 2000 to January 1st, 2019. We consider forecasts at the horizons of the next Fiscal Year (**FY1**) and Second Fiscal Year (**FY2**). Some analysts have multiple revisions during this period. We only consider a revision if it was recorded at least 6 months (or 12 months) before the forecast period end date for the next Fiscal Year (Second Fiscal Year). We use the data from January 1st, 2000 to January 1st, 2012 for training, data from January 1st, 2012 to January 1st, 2014 for validation and data from January 1st, 2014 to January 1st, 2019 for testing.

5 Results

We consider the difference between the actual reported change in EPS and forecasted change using our method and each of the reference baselines. We report the micro averaged Root Mean Squared Error (**RMSE**), the Mean absolute Error (**MAE**) and the Coefficient of Determination ($\boldsymbol{R^2}$) across all companies for Fiscal Year 1 in Table 1 and Fiscal Year 2 in Table 2. We also report the **95%-CI** which we generate by bootstrapping the inferred results for the test data points 1000 times. For completeness, we also report the **RMSE** and **MAE** values Macro Averaged over the individual companies.

	MACRO AVERAGE		MICRO AVERAGE		
MODEL	**RMSE**	**MAE**	**RMSE**	**MAE**	$\boldsymbol{R^2}$
NA	0.4872 ± 0.001	0.4104 ± 0.001	0.7076 ± 0.001	0.4093 ± 0.001	0.7722 ± 0.002
WA	0.5062 ± 0.001	0.4271 ± 0.001	0.7287 ± 0.001	0.4252 ± 0.001	0.7620 ± 0.002
RA	0.5044 ± 0.001	0.4286 ± 0.001	0.7257 ± 0.001	0.4277 ± 0.001	0.7647 ± 0.001
BA	0.9940 ± 0.002	0.8536 ± 0.002	1.5714 ± 0.005	0.8432 ± 0.001	$*$
LBA	0.4809 ± 0.001	0.4039 ± 0.001	0.6980 ± 0.001	0.4029 ± 0.001	0.7809 ± 0.002

Table 1: Results on the Fiscal Year 1 (**FY1**) Year Estimates

	MACRO AVERAGE		MICRO AVERAGE		
MODEL	**RMSE**	**MAE**	**RMSE**	**MAE**	$\boldsymbol{R^2}$
NA	1.0150 ± 0.002	0.8620 ± 0.002	1.6251 ± 0.005	0.8553 ± 0.001	0.4658 ± 0.002
WA	1.0140 ± 0.002	0.8595 ± 0.002	1.6246 ± 0.005	0.8513 ± 0.001	0.4696 ± 0.002
RA	1.0920 ± 0.002	0.9377 ± 0.002	1.7267 ± 0.005	0.9309 ± 0.002	0.3986 ± 0.002
BA	1.2831 ± 0.002	1.1206 ± 0.002	1.7698 ± 0.004	1.1098 ± 0.001	$*$
LBA	1.0055 ± 0.002	0.8494 ± 0.002	1.5877 ± 0.005	0.8386 ± 0.001	0.4837 ± 0.002

Table 2: Results on the Fiscal Year 2 (**FY2**) Year Estimates

From the results in Tables 1 and 2, it is evident that analysts are reasonably accurate in their predictions of future earnings, as evidenced by the low **RMSE** for the unadjusted consensus estimates (**NA**). However, as expected, we see that analysts tend to err more as the forecast horizon is increased, as is evident from the higher **FY2** errors. Although Weighted Averaging (**WA**) reduces errors in the **FY2** consensus estimates, this benefit is not significant given the large confidence intervals around the results. Interestingly, we observed that Regression Adjustment (**RA**) reduced the consensus error by a large margin on the training dataset, but had worse performance on the test set. This was true for both Parametric Ridge Regression and Non-Parametric Random Forest Regression, suggesting that these models have a large tendency to overfit. Furthermore, amongst all the proposed baselines, Bayesian Adjustment **BA** has the highest errors (we do not report the $\boldsymbol{R^2}$ for **BA** for this reason). We hypothesize that this is because Bayesian Adjustment does not allow for the flexibility of discovering analysts who are *unbiased*. In contrast, our proposed Latent Bayesian Adjustment (**LBA**) reduces forecast error across all reported metrics for both **FY1** and **FY2** and the reductions are significant in each case, demonstrating its effectiveness as an improved consensus model.

6 Discussion and Future Work

Biases and systematic errors in earnings forecasts can negatively impact both investors and public companies. Accurate, unbiased consensus earnings are important to investors to understand the financial health of companies and value of their stock so they can make well-informed investment decisions. Similarly, if analysts' estimates are affected by behavioral, incentive-based or discriminatory biases, this may result in companies' stocks being undervalued. We proposed a Bayesian latent variable model and inverse inference procedure that we demonstrated produces estimates which are more robust than consensus estimates as well as other adjusted baselines.

There are a number of possible directions to pursue to further improve the model. Research has shown that analysts whose buy and sell recommendations are more profitable also produce more accurate estimates [Loh and Mianc, 2006]. Adding analysts' recommendations to the model might result in more robust identification of latent subgroups and more accurate estimates. Additionally, while our model incorporates the asymmetry in systematic errors for profit-making and loss-making companies, it does not incorporate other specific biases and asymmetries that have been identified, such as effects for different types of companies, investment banking relationships and discriminatory out of group effects. Additional data identifying some of these attributes for individual analysts and companies could further improve the model and make it more robust to these types of forecasting errors. Furthermore, the model we proposed is linear. Given the observed asymmetries in analysts' systematic forecasting errors, a nonlinear model might further improve the estimation procedure.

While the focus of this paper is generating robust consensus earnings estimates, we note that the proposed model is applicable to any other problem where we have a quantity that is measured by multiple instruments or individuals, which may be subject to machine error or human subjectivity. There are many other close applications in finance and economics, such as GDP and unemployment forecasting, where this model may prove to be more robust than existing approaches. It might also prove useful in more distant applications like elections forecasting or combining sensor readings.

Disclaimer

References

David Blei, Alp Kucukelbir, and Jon McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017.

Mark Thomas Bradshaw. Analysts' forecasts: What do we know after decades of work? *SSRN Electronic Journal*, 2011. doi: 10.2139/ssrn.1880339.

Lawrence D. Brown. Analyst forecasting errors: Additional evidence. *Financial Analysts Journal*, 53(6):81–88, 1997.

Lawrence D. Brown and Michael S. Rozeff. The superiority of analyst forecasts as measures of expectations: Evidence from earnings. *Journal of Finance*, 33(1):1–16, 1978.

Lawrence D. Brown, Robert L. Hagerman, Paul A. Griffin, and Mark E. Zmijewski. Security analyst superiority relative to univariate time-series models in forecasting quarterly earnings. *Journal of Accounting and Economics*, 9(1):61–87, 1987.

Edilberto Cepeda and Dani Gamerman. Bayesian modeling of variance heterogeneity in normal regression models. *Brazilian Journal of Probability and Statistics*, 14(1):207–221, 2000.

Edilberto Cepeda and Dani Gamerman. Bayesian methodology for modeling parameters in the two parameter exponential family. *Revista Estadística*, 57(168-169):93–105, 2005.

Stephen J. Ciccone. Trends in analyst earnings forecast properties. *International Review of Financial Analysis*, 14(1):1–22, 2005.

Werner F. M. De Bondt and Richard H. Thaler. Do security analysts overreact? *The American Economic Review*, 80(2):52–57, 1990.

Michael J. Eames and Steven M. Glover. Earnings predictability and the direction of analysts' earnings forecast errors. *The Accounting Review*, 78(3):707–724, 2003.

John C. Easterwood and Stacey R. Nutt. Inefficiency in analysts' earnings forecasts: Systematic misreaction or systematic optimism? *The Journal of Finance*, 54(4):1777–1797, 1999.

John A. Elliott, Donna R. Philbrick, and Cristine I. Wiedman. Evidence from archival data on the relation between security analysts' forecast errors and prior forecast revisions. *Contemporary Accounting Research*, 12(2):919–938, 2010.

Jennifer Francis and Donna Philbrick. Analysts' decisions as products of a multi-task environment. *The Journal of Accounting Research*, 31(2):216–230, 1993.

Dov Fried and Dan Givoly. Financial analysts' forecasts of earnings: a better surrogate for market expectations. *Journal of Accounting and Economics*, 4(2):85–107, 1982.

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

Harrison Hong, Jeffrey D. Kubik, and Amit Solomon. Security analysts' careeer concerns and herding of earnings forecasts. *RAND Journal of Economics*, 31(1):121–144, 2000.

Sima Jannati, Alok Kumar, Alexandra Niessen-Ruenzi, and Justin Wolfers. In-group bias in financial markets. *SSRN Electronic Journal*, 2019. doi: 10.2139/ssrn.2884218.

Vinesh Jha and Haim Mozes. Creating and profiting from more accurate earnings estimates with starmine professional. *StarMine white paper*, 2001.

Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *Proceedings of the 3rd International Conference on Learning Representations*, 2015.

Roger K. Loh and G. Mujtaba Mianc. Do accurate earnings forecasts facilitate superior investment recommendations? *Journal of Financial Economics*, 80(2):455–483, 2006.

Roni Michaely and Kent L. Womack. Conflict of interest and the credibility of underwriter analyst recommendations. *The Review of Financial Studies*, 12(4):653–686, 1999.

Roni Michaely and Kent L. Womack. Market efficiency and biases in brokerage recommendations. In R. H. Thaler, editor, *Advances in Behavioral Finance, Vol. 2*, chapter 11, pages 389–419. Princeton University Press, 2005.

Roni Michaely, Amir Rubin, Dan Segal, and Alexander Vedrashko. Lured by the consensus: The implications of treating all analysts as equal. *SSRN Electronic Journal*, 2018. doi: 10.2139/ssrn.3128734.

Michael B. Mikhail, Beverly R. Walther, and Richard H. Willis. Does forecast accuracy matter to security analysts? *The Accounting Review*, 74:185–200, 1999.

Patricia C. O' Brien, Maureen F. McNichols, and Hsiou-Wei Lin. Analyst impartiality and investment banking relationships. *The Journal of Accounting Research*, 43(4):623–650, 2005.

Michele O'Neill, Minsup Song, and Judith Swisher. How does prior information affect analyst forecast herding? *Academy of Accounting and Financial Studies Journal*, 15:105–128, 01 2011.

Jana Smith Raedy, Philip Shane, and Yanhua Yang. Horizon-dependent underreaction in financial analysts' earnings forecasts. *Contemporary Accounting Research*, 23(1):291–322, 2006.

Scott A. Richardson, Siew Hong Teoh, and Peter D. Wysocki. Tracking analysts' forecasts over the annual earnings horizon: Are analysts' forecasts optimistic or pessimistic? *SSRN Electronic Journal*, 1999. doi: 10.2139/ssrn.168191.