

And Pythia said: “*Buy not sell*”; An analysis of analyst recommendations betting on sparsity

Veni Arakelian^{a,1}, Mike G. Tsionas^b

^a*Economic Research and Investment Strategy, Piraeus Bank, Vas. Sofias 94 & Kerasountos 1, Athens, Greece, & UCL Centre for Blockchain Technologies.*

^b*Montpellier Business School Université de Montpellier, Montpellier Research in Management & Lancaster University Management School, LA1 4YX, UK.*

Abstract

The goal of this paper is to identify influential analysts who generate abnormal returns when issuing a new recommendation by building a model that bets on the sparsity of typical analysts' recommendation data. Based on Bayesian techniques, we estimate a regression model for the abnormal returns in conjunction with a time-varying Markov switching model for the analysts' recommendations, using the α -stable distribution as prior, and we find that the influential analysts are very few. Additionally, we identify publicly available information that contributes to the abnormal returns besides the analysts' recommendations. Moreover, we study the analysts' herding behavior as an application to exemplify the merits of our method. Our findings show that analysts' herding behavior is not pervasive when the model accounts for the deviation of the analysts' recommendations from the prevailing consensus. Finally, we show that our model performs better than LASSO, elastic net, and the horseshoe prior, both in-sample and out-of-sample.

Keywords: Bayesian; Herding; Horseshoe prior; One-sided stable distributions; Shrinkage methods

JEL: G11; G14; G24

*Corresponding author.

Email addresses: arakelianv@piraeusbank.gr (Veni Arakelian), m.tsionas@lancaster.ac.uk (Mike G. Tsionas)

¹Disclaimer: The views expressed are those of the author and do not necessarily reflect those of the Piraeus Bank.

1. Introduction

Analysts' recommendations literature dates back almost 30 years². A popular question in this literature is whether an abnormal return is generated after a new recommendation. The answer seems to be subject to the model used. For example, Womack [40], Barber et al. [3], Green [22], and most recently, Crawford et al. [16] and Crane and Crotty [15] show that analysts' recommendations generate returns. Conversely, Altinkılıç et al. [1] find that the analysts' forecast revisions announcements release little new information, on average, when intraday returns are used. Further cross-sectional evidence from returns around the announcements confirms that the revisions are virtually information-free. Kim and Song [28] demonstrate that when piggybacking is controlled, the temporal differences in price reactions to analyst forecast revisions disappear.

Analysts' herding behavior is another critical question, though most of the studies illustrate the possibility of herding rather than test it. Graham [21] attempts to identify herding among analysts empirically, showing that analysts' herding behavior is driven by analysts' ability, current reputation status, and prior public information strength. Same, Welch [39] shows that analysts' choices are correlated with the prevailing consensus forecast since there is significant evidence of herding toward the consensus among analysts, which is not related to the consensus accuracy. Thus, analysts do not tend toward the consensus based on fundamental information. Jegadeesh and Kim [27] develop and implement a test to investigate whether sell-side analysts herd around the consensus. Their empirical results support the herding hypothesis. However, some researchers question the herding behavior findings, pointing out that the use of shared information is not adequately controlled, which might be the reason for the clustered recommendations and not the herding behavior toward consensus (e.g., Bernhardt et al. [5]). The conceptualization of herding among analysts is crucial since it might inflate institutional investors' herding consequences. For example, Brown et al. [10] document that mutual fund managers herd, as they massively buy stocks with consensus sell-side analyst upgrades and sell

²Brauer and Wiersema [8] provide a review addressing the questions and the key takeaways.

stocks with consensus downgrades. On the other hand, keeping a considerable distance from the consensus is not a clear signal. Ashiya [2] documents the analyst's incentive to generate publicity by making an extreme recommendation.

The analysts' recommendations data shows a variety of peculiarities. The updates are not scheduled, the recommendations may remain unchanged for a long time (Bernhardt et al. [4]), the stock coverage can be abruptly interrupted, and changes in the analysts' recommendations data provider records are possible (Ljungqvist et al. [29]). Moreover, the recommendations are categorical, contrary to the rest of the variables. Besides, since only a few analysts issue recommendations for a stock on any given day and the rest say nothing, it is obviously a very sparse model. One popular way to deal with it is to assume that only a handful of strong predictors exist at any point in time, that is, to bet on sparsity (Hastie et al. [23]). This assumption can be leveraged by a researcher using a shrinkage estimation method such as LASSO (Tibshirani [34]) or elastic net (Zou and Hastie [42]). To deal with sparse Bayesian estimation, Carvalho et al. [11] introduced the horseshoe prior, which has been found empirically to outperform LASSO but also Bayesian model averaging in terms of out-of-sample predictive sum-of-squares error. Though, the horseshoe prior has its own drawbacks.

In this paper, we propose a model for abnormal returns that uses as regressors each analyst's information (new and immediately preceding recommendation, target price expected return, target price revision, and deviation of the analyst's new recommendation from the prevailing consensus) and variables that capture the market conditions. We employ Bayesian techniques to estimate the model for the abnormal returns in conjunction with a time-varying Markov switching model for the analysts' recommendations. To deal with sparsity, we use the one-sided (or half-stable) distribution as a prior, which imposes heavy zero constraints, allowing the less important coefficients to shrink to zero and, as a result, revealing the influential analysts. As with Chincó et al. [12], we differ from the existing literature by not including variables that other researchers have already identified using the standard econometric methods. Furthermore, we benchmark our model's performance against three alternatives: LASSO, elastic net, and the horseshoe,

and we conclude that it performs better, both in-sample and out-of-sample. We show that very few analysts are influential, so their recommendations have investment value. Additionally, we examine whether analysts herd by distinguishing between *conditional* and *unconditional herding*. Conditional (unconditional) herding is inferred by the relationships among the analysts' recommendations coefficients obtained from the estimation of the abnormal returns model when we account (do not account) for the effect of the deviation of the analyst's new recommendation from the prevailing consensus. We show that conditional herding is not pervasive given the sparse structure of the coefficient estimates. On the contrary, the structure is very dense when unconditional herding is assumed.

The structure of the paper is as follows: Section 2 describes sparse Bayesian estimation and goes over some prior formulations used to model sparsity. Section 3 details our data. Section 4 develops the model. Section 5 presents the empirical results. Section 6 evaluates the performance of our proposed methodology, and section 7 concludes the paper.

2. Sparse Bayesian Estimation

Buy and hold abnormal returns is a standard method of measuring long-term abnormal returns. Jegadeesh and Kim [27] compute the h -day buy-and-hold abnormal returns, $ABR_{x,t,t+h}$, after a recommendation revision for stock x on date t , as $ABR_{x,t,t+h} = \prod_{\tau=t}^{t+h} (1 + R_{x,\tau}) - \prod_{\tau=t}^{t+h} (1 + R_{m,\tau})$, where $R_{x,\tau}$ is the daily return of the stock x , and $R_{m,t}$ is the market index return. Day 0 is the recommendation date or the next trading day. Assume that $ABR = X\beta + v$ is a model for the ABR, where $X \in \mathbb{R}^{n \times K}$ is the explanatory variables matrix, $\beta \in \mathbb{R}^K$ is the vector of unknown parameters, and v is the error term. Often the number of regressors (K) is large or too large for us to interpret them individually. We typically want to minimize the sum of squares plus some penalty on the coefficients. In some instances, we may also have $K \gg n$ and are interested in exploiting sparsity because many regression coefficients will be zero. To do so, regularization methods were developed. For example, the formulation $\min_{\beta \in \mathbb{R}^K} (ABR - X\beta)'(ABR - X\beta) + \lambda \sum_{k=1}^K |\beta_k|$, $\lambda \geq 0$ leads to the LASSO estimator (Tibshirani [34]). The “compromise”

objective function $\min_{\beta \in \mathbb{R}^K} (ABR - X\beta)'(ABR - X\beta) + \lambda_1 \sum_{k=1}^K |\beta_k| + \lambda_2 \sum_{k=1}^K \beta_k^2$, $\lambda_1, \lambda_2 \geq 0$ is called elastic net and it was developed to cope with the limitations of LASSO (Zou and Hastie [42]). λ (or λ_1, λ_2 for the elastic net) is the penalty parameter. λ is typically chosen either using information criteria, cross-validation, or a Bayesian prior, which rests on the idea that the LASSO is connected to the Laplace distribution which has a mixture interpretation.

To deal with sparse Bayesian estimation, Carvalho et al. [11] introduced the horseshoe prior as a global-local Gaussian scale mixture under a half-Cauchy prior with density $p(\beta_k|\tau) = \int_0^\infty \frac{1}{\tau\lambda_k} \phi\left(\frac{\beta_k}{\tau\lambda_k}\right) \frac{2}{\pi(1+\lambda_k^2)} d\lambda_k$, $\tau > 0$, and $\phi(\cdot)$ is the standard normal density. In particular, for the model $y|\beta \sim \mathcal{N}(\beta, \sigma^2 I)$, when it is believed that β is sparse, Carvalho et al. [11] proposed to use the following specification

$$\beta_k|\lambda_k \sim \mathcal{N}(0, \lambda_k^2), \lambda_k|\tau \sim \mathcal{C}^+(0, \tau), \tau|\sigma \sim \mathcal{C}^+(0, \sigma), \quad (1)$$

where $\mathcal{C}^+(0, s)$ is the half-Cauchy distribution with zero location and scale parameter s defined over positive real numbers. This horseshoe prior has no tuning parameters and has excellent performance. The half-Cauchy distribution is one of the Cauchy distribution's symmetric halves (if unspecified, it is the right half that's intended). The half-Cauchy has many properties; some of these properties are useful, and we may want them in the prior. A common choice for a prior on a scale parameter is the inverse gamma (not least, because it's conjugate for some familiar cases). When a weakly informative prior is desired, very small parameter values are used. The half-Cauchy is quite heavy-tailed, and it, too, might be regarded as fairly weakly informative in some situations. Gelman [19] advocates for half-t priors (including the half-Cauchy) over the inverse gamma because they have better behavior for small parameter values but only regard it as weakly informative when a large scale parameter is used. Gelman has focused more on the half-Cauchy in more recent years. The paper by Carvalho et al. [11] gives additional reasons for choosing the half-Cauchy in particular. For σ , the prior $p(\sigma) \propto \sigma^{-1}$ is assumed.

Such global-local priors have been introduced by Carvalho et al. [11], Polson and Scott [32],

and Polson and Scott [33] to account for sparsity via global shrinkage and identify parameters by local shrinkage. Many authors (Polson and Scott [32], Polson and Scott [33], Datta and Ghosh [18]) have found empirically that horseshoe outperforms LASSO but also Bayesian model averaging in terms of out-of-sample predictive sum-of-squares error. “[I]t has a closed form marginal prior for β , yet with a spike at origin and heavy tails and more importantly, admits a global–local scale mixture representation” (Bhadra et al. [6], p. 412). Moreover, horseshoe and other so-called global–local priors are optimal in variable selection, estimation, and forecasting. Horseshoe estimators are important since they have near-minimax rates in both an empirical Bayes and full Bayes approach (van der Pas et al. [37], van der Pas et al. [38], van der Pas et al. [35], van der Pas et al. [36]). The horseshoe prior has proven to be a noteworthy alternative for sparse Bayesian estimation, but has previously suffered from two problems. First, there has been no systematic way of specifying a prior for the global shrinkage hyperparameter based on prior information about the degree of sparsity in the parameter vector. Second, the horseshoe prior has the undesired property; there is no possibility of specifying separately information about sparsity and the amount of regularization for the largest coefficients, which can be problematic with weakly identified parameters, such as the logistic regression coefficients in the case of data separation.

2.1. Horseshoe-like Priors

A proper prior that has the same behavior is the horseshoe-like prior (Bhadra et al. [6]), which has the following form: $p(\beta_k|\tau) = \frac{1}{2\pi\tau} \log\left(1 + \frac{\tau^2}{\beta_k^2}\right)$, $\beta_k \in \mathbb{R}, k = 1, \dots, K$, $\tau > 0$. Setting $\alpha = \tau^2$, it turns out that this can be written as

$$\beta_k|\lambda_k, a \sim \mathcal{N}\left(0, \frac{a}{2\lambda_k}\right), p(\lambda_k) = \frac{1}{2\sqrt{\pi}}\lambda_k^{-3/2}(1 - e^{-\lambda_k}), a > 0, \lambda_k > 0, k = 1, \dots, K.$$

If $\mathbb{E}(u_i^2) = 1$, then the posterior mode is given by $\mathbb{E}(\beta|a, y, \lambda, X) = [X'X + D(\lambda)]^{-1} X'y$, where $\lambda = [\lambda_1, \dots, \lambda_K]'$, $D(\lambda) = \text{diag}\left(\frac{2\lambda_k}{a}\right)$, and $(X'X + D^{-1})^{-1} = D - DX'(XDX' + I_n)^{-1}XD$, which is $O(nK^2)$ for large K . One can update these elements via an EM algorithm using for

λ the form $\lambda_k = \frac{1}{2\pi a^{1/2}} \left(\frac{a}{\beta_k^2} - \frac{a}{a+\beta_k^2} \right)$, $k = 1, \dots, K$, and for a , the updating scheme, $a := \frac{a^{3/2}}{K\pi} \sum_{k=1}^K \frac{1}{a+\beta_k^2}$. By Bhadra et al. [6], Lemma 5.1, we use the following representation for Gibbs sampling:

$$\beta_k | t_k, \tau \sim \mathcal{N} \left(0, \frac{\tau^2}{t_k^2} \right), t_k | s_k \sim \mathcal{N}(0, s_k), s_k \sim \mathcal{P} \left(\frac{1}{2} \right),$$

where $t_k^2 = 2\lambda_k$, $\tau > 0$, $s_k > 0$, $k = 1, \dots, K$, and \mathcal{P} denotes the Pareto distribution.

One-sided (or half-stable) distribution can be used instead of the horseshoe-like prior. The distribution function of the half-stable distribution is denoted by $\mathcal{S}(\alpha, \beta, d, c)$, where d, c are location and scale parameters, $0 < \alpha \leq 2$ is the characteristic exponent, and $-1 \leq \beta \leq 1$ is the skewness parameter. For $d = 0, s = 1$ we have the standard form. When $\beta = 1$, the distribution is ‘‘maximally skewed’’, and for $\alpha < 1$, it is defined over positive reals. The distribution obtained is an α -stable distribution, and our interest focuses on this. The family is denoted by $\mathcal{S}_\alpha(c)$, and its density is denoted by $f(x; \alpha, c)$. So, we can replace (1) by

$$\beta_k | \lambda_k \sim \mathcal{N}(0, \lambda_k^2), \lambda_k | c \sim \mathcal{S}(\alpha, 1, 0, c), c | \sigma \sim \mathcal{S}(\alpha, 1, 0, \sigma), \alpha < 1, k = 1, \dots, K. \quad (2)$$

The joint prior of the regression coefficients from (2) is as follows.

$$p(\boldsymbol{\beta}, \lambda, \alpha, c) = \left\{ \prod_{k=1}^K (2\pi\lambda_k^2)^{-1/2} e^{-\beta_k^2(2/\lambda_k^2)} f(\lambda_k; \alpha, c) \right\} f(c; \alpha, \sigma)^K, \quad (3)$$

where $f(\lambda_k; \alpha, c)$ is the density of $\mathcal{S}_\alpha(c)$, $f(c; \alpha, \sigma)$ is the density of $\mathcal{S}_\alpha(\sigma)$, and $\lambda = [\lambda_1, \dots, \lambda_K]'$. Conditionally on λ , drawing β involves drawing the parameters $\boldsymbol{\beta}$ from a heteroskedastic linear model. The conditional posterior of each λ_k is

$$p(\lambda_k | \beta_k, \alpha, c) \propto \lambda_k^{-1} e^{-\beta_k^2(2/\lambda_k^2)} f(\lambda_k; \alpha, c), k = 1, \dots, K. \quad (4)$$

We draw a candidate $\lambda_k^{(c)}$ from a distribution whose density is $q(\lambda_k) \propto \lambda_k^{-(\bar{\nu}+1)} e^{-\beta_k^2(2/\lambda_k^2)}$ (Zellner [41], p. 371, equation A37b) where $\bar{\nu}$ is a parameter. The candidate is accepted with probability

$\min \left\{ 1, \frac{p(\lambda_k^{(c)} | \beta_k, \alpha, c) / q(\lambda_k^{(c)})}{p(\lambda_k^{(s)} | \beta_k, \alpha, c) / q(\lambda_k^{(s)})} \right\}$, where s denotes the current number of MCMC iteration. We select the parameter $\bar{\nu}$ using some initial experimentation with representative values of β_k , α and c and, finally, select $\bar{\nu} = 4$ yielding acceptance rates close to 30% on average. The conditional posterior of c is

$$p(c | \alpha, \sigma)(c; \alpha, \sigma)^K \prod_{k=1}^K f(\lambda_k; \alpha, c). \quad (5)$$

We use inversion of the numerical cumulative distribution function to provide an “exact” draw from this conditional posterior using 50 points in its support. This involves the computation of a stable density function but the computation at the 50 different points is easily parallelizable. We use the same technique to generate random numbers from $p(\sigma | \alpha, c) \propto f(c; \alpha, \sigma)^K$ and the conditional posterior of the characteristic exponent, viz. $p(\alpha | \lambda, c) \propto p(\alpha) \prod_{k=1}^K f(\lambda_k; \alpha, c)$, where $p(\alpha)$ denotes a flat prior on α .

3. Data Description

We obtain 925,555 analysts’ recommendations from I/B/E/S for 4,173 stocks, which are current or past constituents of the S&P 500, Nasdaq, and NYSE indices, issued by 12,716 analysts (or 833 brokerage firms) identified through their analyst code³. The sample covers the period from January 1, 1999, to April 30, 2020. I/B/E/S transforms the analysts’ recommendations (e.g. “buy (hold, sell)”, “market underperform (outperform)”, etc) to numerical scores where “5” stands for strong sell, “4” for sell, “3” for hold, “2” for buy and “1” for strong buy. If the analyst has no opinion on what to recommend, a “0” recommendation is issued. In our data set, “no opinion” recommendations account for 0.1% of the recommendations will be excluded from our analysis. From 837,107 recommendations, excluding the initial and “no opinion” recommendations, 108,972 are upgrades, 103,866 are downgrades, and 624,269 are reiterations, confirming the

³There are additionally 420 analysts who are not identifiable. Additionally, our sample satisfies the criteria mentioned by other studies, (e.g., Jegadeesh and Kim [27]) that is, more than two analysts issue recommendations in the same period for a stock whose price should be at least \$1 on the day before the recommendation revision date. The interested reader can find the breakdown per year of the number of recommendations, recommendation changes, recommendations per level, analysts, and brokerage firms in Online Appendix, Table 1.

reluctance of the analysts to change their recommendations⁴ (Bernhardt et al. [4]). Furthermore, only 6,571 of the recommendation changes are three- and four-class upgrades/downgrades, and contrary to some studies, we keep them in our analysis.

The majority of analysts' recommendations are accompanied by a target price for the stock, which is very important for investors (Huang et al. [26]). Based on the target price, we model analyst's exaggeration through the analyst's *target price expected return* (Huang et al. [25]), and the analyst's *target price revision* (Brav and Lehavy [9]). The analyst's target price expected return (*TER*) is the return between the price of the stock at the date before the recommendation announcement and the analyst's target price. The analyst's target price revision (*DTar*) is the return between the analyst's new target price and his previous one. Following Jegadeesh and Kim [27], we also employ the deviation of the analyst's new recommendation from the prevailing consensus (*NC*). We find that the distributions are very skewed with very long right tails showing that the analysts are very optimistic⁵.

To calculate the abnormal returns, we use the daily price changes of the stock and the market index (i.e., S&P 500, Nasdaq or NYSE Index) quoted by Refinitiv Eikon. In addition, we include variables to capture the market conditions that might affect the abnormal returns directly, and indirectly the analyst's recommendation (Conrad et al. [14], Peng and Xiong [31]). The variables are the following: 1. VIX Index, 2. Price to earnings (P/E) of S&P 500 Index, 3. Growth rate of the trailing 12-month sums of earnings of S&P 500 Index, 4. DEBT/EBITDA ratio of the S&P 500 Index, 5. Net equity expansion on the NYSE Index, calculated as $NetIssue_t = Mcap_t - Mcap_{t-1}(1 + vwretx_t)$, where $Mcap$ is the total market capitalization, and $vwretx$ is the value weighted return (excluding dividends) on the NYSE Index, 6. Yield to maturity for the three-month Treasury bill, 7. 2/10 U.S. Government bond yield spread, 8. Moody's BAA and AAA-rated corporate bond yields spread, 9. U.S. inflation, 10. U.S. ISM Manufacturing Purchasing Managers Index, 11. Growth rate of the U.S. Conference Board

⁴The sample transition probability matrix is given in Online Appendix, Table 3.

⁵See Online Appendix, Table 4.

Leading Economic Indicators Index, 12. U.S. Business Cycle Phase, suggesting the four stages of the business cycle: expansion, slowdown, contraction, and recovery. Similar variables are used to construct the investor sentiment to study its correlation with analysts' forecast errors (Hribar and McInnis [24]).

4. Model

Jegadeesh and Kim [27] to determine if the analyst's recommendations generate abnormal returns fit the following regression model,

$$ABR_{x,t,t+h} = \alpha_h + b_h \times I_{multi} + c_h \times I_{single} + d_h \times NC_{x,t} + v_{t,h}, \quad (6)$$

where $h = \{0, 1, 2\}$ is the event window, and $v_{t,h} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_h^2)$ the error term. The dummy variables I_{multi} and I_{single} take the value of 1, if there is a multi/single upgrade, and the value of -1, if there is a multi/single downgrade. I_{single} takes the value of zero if there is no recommendation change. Obviously, they take the sign of expected abnormal returns conditional on an upgrade or a downgrade, and they are used to pool upgrades and downgrades of different level changes in the same regression. Additionally, according to Jegadeesh and Kim [27], if analysts herd close to the consensus, then $d > 0$. On the contrary, when analysts exaggerate their differences with the consensus, then $d < 0$.

In the spirit of (6), we build a model for the abnormal returns accounting for the each analyst's recommendations and the regressors discussed in Section 3. Regarding the analyst's recommendations, we introduce the dummy variables $\mathbf{I}(R_{t-1} = i)$ and $\mathbf{I}(R_t = j)$, that take the value 1, if the analyst's recommendation is i ($j \in \{-2, -1, 0, 1, 2\}$) at $t - 1$ (t). The past recommendation accounts for the fact that "prior analyst's choices are an important influence on the next recommendations" (Welch [39], p. 370). The analyst's view about the stock is

completed with the variables TER , $DTar$, and NC . Therefore, the full model is:

$$ABR_{x,t,t+h} = \mathbf{X}'_t \boldsymbol{\beta}_1 + \left(\underbrace{TER_t, DTar_t, NC_t, \mathbf{I}_{R_{t-1}}, \mathbf{I}_{R_t}}_{(1)} \dots \underbrace{TER_t, DTar_t, NC_t, \mathbf{I}_{R_{t-1}}, \mathbf{I}_{R_t}}_{(G)} \right) \boldsymbol{\beta}_2 + v_{t+h} \quad (7)$$

where $\mathbf{X}_t \in \mathcal{X}$ contains the market conditions variables discussed in Section 3, \mathbf{w}_t contains the variables of all the G analysts covering the stock, and $\boldsymbol{\beta}_2$ is a conformable parameter vector. Along these lines, we can determine (i) whose analysts' recommendations generate abnormal returns, (ii) whether a given analyst follows the recommendations of *other* analysts, and (iii) which market conditions variables are important in a given horizon $h = 0, 1, 2$.

Next, we assume that analyst's recommendations, R_t and R_{t-1} , are governed by a time-varying Markov switching model with transition probabilities mapping the recommendations "strong sell", "sell", "hold", "buy", and "strong buy" to the numbers -2,-1,0,1,2, respectively. The analyst moves from his previous recommendation (row-) i to his new recommendation (column-) j , with probability $p_{ij,t} = Pr(R_t = j | R_{t-1} = i)$, $i, j \in \{-2, -1, 0, 1, 2\}$. The transition probability matrix \mathbf{P} is normalized so that the sum in each row is 1. Suppose that \mathbf{z}_t includes the variables, TER , $DTar$, and NC . We propose to parameterize the transition probabilities as

$$p_{ij,t} = \frac{\Phi(z'_{it} \gamma_{ij})}{\sum_{m=1}^M \Phi(z'_{it} \gamma_{im})}, t = 1, 2 \dots, T, \quad (8)$$

where γ_{ij} , $i, j \in \{-2, -1, 0, 1, 2\}$, are coefficients arranged so that the intercepts corresponding to each equation are in increasing order, $\Phi(\cdot)$ is the standard normal distribution function, and, M is the number of states, which in our case is five. Additionally, we modify the distributional assumption in (6) as $v_{t,h} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the variance-covariance matrix of the error terms. To model endogeneity, we include \mathbf{v}_t and \mathbf{v}_{t-1} among the regressors \mathbf{z}_t that determine the transition probabilities. Computationally, this is not particularly demanding as we use a fast MCMC to provide statistical inferences⁶. Moreover, the α -stable distribution prior comes

⁶See, Online Appendix, Section 2.

in handy as we have a large number of parameters.

If $\boldsymbol{\theta} \in \Theta$ denotes the entire vector of parameters, the likelihood function for date t is

$$l_t(\boldsymbol{\theta}) = |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{U}_t' \boldsymbol{\Sigma}^{-1} \mathbf{U}_t \right\} \Pr(R_t = j, R_{t-1} = i), \quad (9)$$

where $\mathbf{U}_t = [ABR_{x,t,t+h} - \mathbf{X}_t' \boldsymbol{\beta}_1 - \mathbf{w}_t \boldsymbol{\beta}_2, h \in \mathbb{H}]'$ assuming $R_{t-1} = i$ and $R_t = j$, $i, j \in \{-2, -1, 0, 1, 2\}$. The probability $\Pr(R_t = j, R_{t-1} = i)$ is defined by

$$\Pr(R_t = j, R_{t-1} = i) = \sum_i \Pr(R_t = j | R_{t-1} = i) \Pr(R_{t-1} = i). \quad (10)$$

The overall likelihood is $L(\boldsymbol{\theta}) = \prod_{t=1}^T l_t(\boldsymbol{\theta})$. Apart from drawing $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2), \lambda, \alpha, c, \sigma$, the remaining parameters are drawn using a fast MCMC procedure as described in the Online Appendix, Section 2.

5. Results

The results are inferred from LASSO, elastic net, the horseshoe and the α -stable distribution prior applied to the analysts' recommendations data⁷.

5.1. Analyst's Recommendation Transition Probability Mechanism

Table 1 contains the transition probabilities (8) evaluated at the mean of the posterior distributions of the z_t coefficients, assuming the α -stable distribution prior⁸. The transition probabilities notably differ from their empirical counterparts⁹. Analysts are more reluctant to change “sell” (41%) or “buy” (30.2%) rather than “hold” (21.4%), “strong sell” (16.1%), or “strong buy” (10.3%) recommendations. They are very keen to do two-level upgrades (all statistically significant at a 1% significance level), or downgrade a stock from “strong buy” to “hold” (24.4%)

⁷Additionally, we run the exercise using the brokerage recommendations data. Selected results are reported in the Online Appendix, Section 3.

⁸The posterior distribution of α has a mass between 0.58-0.72, suggesting that the distribution is indeed a “maximally skewed” stable distribution. See also, Online Appendix, Figure 1.

⁹See Online Appendix, Table 3.

rather than to “strong sell,” “sell,” or “buy,” while there is a weak evidence of downgrades from “buy” to “sell” recommendation. This is due to the unwillingness of the analysts to issue negative recommendations, mainly because these firms are accounted for as future investment banking clients. Additionally, the stock might underperform for several reasons, for example, an event causing market stress; therefore, the analyst keeps it in the portfolio in the prospect of over-performance. Furthermore, analysts willingness to change their recommendation from “strong sell” to “buy” (31%), highlights their aversion to negative recommendations again. The three-levels downgrade from “buy” to “strong sell” is statistically significant only at 10%. Finally, the transition probabilities from “hold” recommendations range from 21.4%-27.1%, excluding the upgrades to “buy” recommendations, which are not statistically significant. The re-iteration of “hold” recommendations has the smallest probability (21.4%), suggesting that analysts, through their recommendations, seek to generate fees for the brokerage firms they work for. Lastly, the four-level upgrade/downgrade is statistically significant only at 10%.

Table 1: Posterior mean estimates of analysts’ recommendations transition probabilities

	Strong sell	Sell	Hold	Buy	Strong buy
Strong sell	0.161*** (0.021)	0.272*** (0.014)	0.252*** (0.044)	0.310*** (0.019)	0.005* (0.003)
Sell	0.108 (0.033)	0.410*** (0.045)	0.278*** (0.017)	0.189*** (0.015)	0.015 (0.020)
Hold	0.254*** (0.076)	0.271*** (0.049)	0.214*** (0.055)	0.006 (0.004)	0.255*** (0.048)
Buy	0.108* (0.065)	0.178* (0.098)	0.230*** (0.047)	0.302*** (0.092)	0.182** (0.081)
Strong buy	0.228* (0.133)	0.220 (0.155)	0.244*** (0.016)	0.205*** (0.033)	0.103*** (0.017)

Note: The table reports the sample means (with sample standard deviations in parentheses) of posterior mean estimates across the sample for the transition probabilities assuming $h = 0$. ***, **, and *denote the statistical significant transition probabilities at 1%, 5% and 10% significance level, respectively.

Table 2 summarizes the impact of TER , $DTar$, and NC , to the transition probabilities. The positive coefficient on a covariate means that the stocks with a higher value of that covariate are more likely to move to that recommendation level. We comment only the major contribution

of each variable in absolute value. TER positively affects the probability of reiterating “sell” recommendations (0.1524), and negatively, the downgrade from “buy” to “hold” (-0.2394). Same, $DTar$, positively affects the probability of reiterating “strong sell” recommendation (0.1764), and negatively, the upgrade from “hold” to “buy” (-0.1715). Finally, NC , positively affects the transition probability from “hold” to “strong sell”, and negatively, the downgrade from “buy” to “hold” (-0.0765).

Table 2: Transition probabilities drivers

	Strong sell			Sell			Hold		
	TER	DTar	NC	TER	DTar	NC	TER	DTar	NC
Strong sell	0.0861*** (0.0193)	0.1764*** (0.0382)	0.0169 (0.0142)	0.0376*** (0.0144)	-0.0350 (0.0850)	0.1175*** (0.0133)	0.0284 (0.0299)	0.1183*** (0.0256)	-0.0191** (0.0094)
Sell	0.1401*** (0.0383)	0.0714* (0.0370)	0.0628*** (0.0205)	0.1524*** (0.0398)	-0.0502 (0.0406)	0.0418** (0.0170)	0.0526*** (0.0130)	-0.0531 (0.0714)	0.0076 (0.0101)
Hold	-0.0449 (0.0346)	-0.0162 (0.0491)	0.1980*** (0.0100)	-0.0747** (0.0352)	-0.0412 (0.0862)	0.0994*** (0.0168)	0.1242*** (0.0316)	0.1019* (0.0532)	0.0499** (0.0227)
Buy	0.0720*** (0.0271)	0.0406 (0.0408)	0.0550* (0.0298)	-0.1120*** (0.0200)	-0.0831 (0.0786)	-0.0765*** (0.0185)	-0.2394*** (0.0329)	-0.1375*** (0.0359)	-0.0250* (0.0144)
Strong buy	-0.0301 (0.0254)	-0.0077 (0.0426)	0.1073*** (0.0052)	0.0339 (0.0350)	0.0516 (0.0764)	-0.0049 (0.0176)	0.0867*** (0.0118)	0.1141*** (0.0126)	0.1157*** (0.0133)
		Buy			Strong buy				
		DTar	NC	TER	DTar	NC			
Strong sell	0.0160 (0.0253)	0.1632** (0.0769)	0.0435 (0.0294)	0.003*** (0.001)	0.0017*** (0.0005)	0.034*** (0.011)			
Sell	-0.0290*** (0.0185)	0.0044*** (0.0595)	0.0671 (0.0299)	0.0014*** (0.0003)	-0.0019*** (0.0004)	-0.023*** (0.0001)			
Hold	-0.0625*** (0.0190)	-0.1715*** (0.0414)	-0.0270 (0.0281)	-0.0020*** (0.0004)	0.0044*** (0.0011)	-0.010** (0.004)			
Buy	0.0310 (0.0358)	0.0523 (0.0637)	-0.0113 (0.0080)	-0.004*** (0.001)	-0.0031*** (0.0006)	-0.033*** (0.011)			
Strong buy	0.0279 (0.0246)	0.1063 (0.0895)	-0.0332** (0.0132)	-0.001 (0.001)	0.045*** (0.012)	-0.003*** (0.0005)			

Note: Estimates of the coefficients γ_{ij} obtained from (8). The standard errors of the estimates are reported in the parenthesis. ***, **, and * denote the statistical significant coefficients at 1%, 5% and 10% significance level, respectively.

5.2. Cross-sectional Determinants of Abnormal Returns

Table 3 reports the cross-sectional regression results for the coefficients of (7) using the α -stable distribution prior, at the recommendation revision day ($h = 0$). The coefficient associated with TER is significantly larger than that of the rest variables, confirming that it provides valuable information regarding a firm's value not already reflected in analysts' recommendations (Da et al. [17]). Like in Brav and Lehavy [9], we document significant abnormal returns increasing in the favorableness of $DTar$. It seems that both TER and $DTar$ bring new information to the market that is not subsumed by the analyst's recommendation, capturing the implications of both short-term earnings forecast revisions and P/E ratio revisions, strengthening the argument that target price is informative. Since both embody information about the earnings or the P/E ratio, they highlight the analysts' views about the firm's prospects' overall assessment. In its turn, the positive slope coefficient of NC suggests that as the new recommendation deviates from the consensus, it conveys information to the market, adding to the findings of Jegadeesh and Kim [27]. Abnormal returns are more favorable for upgrades and more damaging for downgrades when the new recommendation is farther away from the consensus than close. The positive herding coefficient supports the hypothesis of herding toward the consensus, and we employ further analysis of this in Section 5.4. Finally, we identify new factors that contribute to abnormal returns. The statistically significant market conditions variables that generate abnormal returns are the S&P 500 P/E Index, the 2/10 U.S. government bond yield spread, the U.S. inflation, and the U.S. ISM Manufacturing Purchasing Managers Index. In contrast to the other three variables that favor index investing, an increase in the S&P 500 P/E Index favors stock investing.

Table 3: Cross-sectional regression results

Variable	Posterior mean estimates
<i>TER</i>	0.038*** (0.014)
<i>DTar</i>	0.012*** (0.003)
<i>NC</i>	0.010*** (0.003)
VIX Index	-0.005 (0.005)
S&P 500 P/E Index	0.017*** (0.004)
Trailing 12M Earnings growth rate of S&P 500 Index	0.005 (0.004)
DEBT/EBITDA of S&P 500 Index	-0.003 (0.004)
Net equity expansion	0.012 (0.005)
3M T-bill yield	0.005 (0.005)
2/10 U.S. Government bond yield spread	-0.008*** (0.002)
Moody's BAA and AAA-rated corporate bond yields	0.003 (0.002)
U.S. inflation	-0.004*** (0.001)
U.S. ISM Manufacturing Purchasing Managers Index	-0.015*** (0.003)
U.S. Conference Board Leading Economic Indicators Index growth rate	0.002 (0.002)
U.S. business cycle phase	0.005 (0.004)
$I(R_{t-1} = -2)$	-0.004** (0.002)
$I(R_{t-1} = -1)$	-0.004*** (0.001)
$I(R_{t-1} = 0)$	0.003*** (0.001)
$I(R_{t-1} = 1)$	0.002** (0.001)
$I(R_{t-1} = 2)$	0.004** (0.002)

Continued on next page

Table 3 – continued from previous page

Variable	Posterior mean estimates
$I(R_t = -2)$	-0.015^{***} (0.003)
$I(R_t = -1)$	-0.0073^{***} (0.001)
$I(R_t = 0)$	0.020^{***} (0.004)
$I(R_t = 1)$	0.015^{***} (0.004)
$I(R_t = 2)$	0.018^{***} (0.005)

Note: The table reports the posterior mean estimates of coefficients of (7). The standard errors of the estimates are reported in the parenthesis. *** , ** , and * denote the statistical significant coefficients at 1%, 5% and 10% significance level, respectively.

Furthermore, we examine the structure of the market conditions coefficients. To have a visualization of the exact relationships, we randomly draw 200 analysts. Additionally, to compare the sparsity among the models and, which will imply later, the percentage of influential analysts who generate abnormal returns, we report the normalized $\|\ell\|_0$ -norm defined by $\|A\|_{norm} := \#\{(i, j), A_{i,j} \neq 0\} / (n * m)$, $A \in \mathbb{R}^{n*m}$. For the non-zero elements, we report the standard descriptive statistics. The estimated coefficients of our analysts' random sample are plotted in Figure 1. The relationship structure emphasizes how the importance of shared information differs. It may justify analysts' favor (or not) over certain stocks, given their out-performance in current market conditions. Analysts' choices might be wrongly interpreted as herding behavior since these variables account for common information among analysts and investors. As Jegadeesh and Kim [27] mention, "analysts may herd or take similar actions because they receive correlated information". Likewise, Graham [21] argues that herding likelihood increases with the level of correlation across informative signals. The coefficient estimates based on the α -stable distribution prior are shown in 1(a). It is shown that the structure is very sparse, except for the U.S. business cycle phase variable, where 140 out of the 200 randomly chosen

analysts assign a positive coefficient, ranging from 0.0007 to 0.2958, which is the highest among all the other market conditions variables. Figures 1(b)-(d) show the coefficients of the market conditions variables using the horseshoe prior, LASSO, and elastic net. The figures are very much alike but much more dense than Figure 1(a). For the horseshoe prior, the least important variable is the growth rate of the trailing 12-month sums of earnings of S&P 500 Index, and the most important one is U.S. inflation (the coefficients are equal to 0.0002 and 0.2977, respectively). Similarly, the yield to maturity of the three-month Treasury bill and the growth rate of the trailing 12-month sums of earnings of the S&P 500 Index are the least and most important variables for LASSO (the coefficients are equal to 0.000 and 0.3283, respectively). Finally, for the elastic net, the least and most important variables are the U.S. business cycle phase and the U.S. ISM Manufacturing Purchasing Managers Index (the coefficients are equal to 0.0004 and 0.3163, respectively). All the figures in Figure 1 acknowledge the crucial role of the U.S. business cycle phase variable by assigning a positive coefficient by more than 132 out of 200 analysts. This is because a sector's, and consequently a stock's, outperformance is closely related to the business cycle phase. For example, a cyclical company follows the overall economy's trends, making its stock price very volatile. So, if the economy enters the recession phase, the analyst might bias his recommendation downwards for a cyclical stock.

5.3. Posterior Distributions of Abnormal Returns

Continuing reporting the results of Table 3, it appears that all the dummy variables are statistically significant, and the signs of the coefficients are in line with what is expected. The coefficients of the “positive” recommendations satisfy the expected ordering (“strong buy” \succ “buy”) and demonstrate that “strong buy” recommendations generate larger price reaction than “buy” recommendations, either t or $t-1$. Similarly, for the “negative” recommendations, “strong sell” recommendations have a more significant negative price impact than “sell” recommendations at time t , and the same effect at $t-1$. “Hold” recommendations have a positive impact on the abnormal returns, implying that “hold” is accounted as a positive recommendation, despite the fact that a savvy investor might interpret it as essentially “sell” (Bradshaw [7]). In Figure

2, for the model handling the prior of β with the α -stable distribution, we plot the marginal posterior densities which do not have a point mass at zero, of the coefficients of $I(R_t = 1)$ for each analyst¹⁰. It is interesting to note that the means of the distributions are far from zero and that analysts' number is dramatically diminished as the days pass, implying that not all analysts' recommendations generate abnormal returns. To conclude, the central question of whether an analyst's new or his previous recommendations have investment value is positively answered. Still, it does not hold for all the analysts covering a stock.

Figure 3(a) shows the average posterior predictive densities of ABR across all analysts¹¹. Table 4 shows that the reactions to recommendation changes yield an average posterior abnormal return of 3.5% on the day of the event, and 2.2% and 1.8% for $h = 1$ or $h = 2$ -days holding period after the recommendation revision, highlighting the cost of investment delay.

Table 4: Posterior moments of abnormal returns

	$h = 0$	$h = 1$	$h = 2$
Mean	0.035	0.022	0.018
Median	0.033	0.023	0.017
StDev	0.0032	0.0014	0.0017
Mean Absolute Deviation	0.0026	0.0011	0.0014

Note: Posterior mean estimates of the abnormal returns $ABR_{x,t,t+h}$ evaluated for $h = 0, 1, 2$ days holding period following the revision date t , when the α -stable distribution prior is used for the estimation (7).

This finding agrees with the existing literature, which documents that while security analysts' stock recommendations lead to an immediate price reaction, a drift continues, adding to the argument of markets' semi-strong inefficiency, where public information can predict future stock returns (see, for example, Womack [40], Barber et al. [3]). If the market reacts efficiently to the information in recommendations, there should not be any predictable drift in stock prices. Figure 3(b) shows the corresponding cumulative distribution functions to examine stochastic dominance. The strategy of the immediate reaction to the analyst's recommendation dominates.

We further investigate whether profitable investment strategies based on analysts' recommen-

¹⁰Similarly, see Online Appendix, Figure 2 for the marginal posterior densities of the $I(R_{t-1} = 1)$ coefficients.

¹¹For each analyst, see Online Appendix, Figure 3.

dations exist. Figure 4 depicts the posterior distributions of expected ABR using the α -stable distribution prior, the horseshoe prior, LASSO and elastic net, the day after the recommendation ($h = 1$), controlling for one of the possible recommendations. All the expected ABR have positive means, implying that analysts' recommendations have investment value even one day after the revision. Figure 4(a) shows that the most profitable strategy is to follow "strong buy" recommendations, and after that, "strong sell", "sell", "hold" and "buy" recommendations. Furthermore, in Figures 4(b)-(d), the posterior densities of the expected ABR of "strong buy" recommendations become bimodal. Still, the mode of the highest return is much bigger than that centered at zero. Our findings are in line, for example, with Womack [40], who show that favorable/unfavorable changes in individual analyst recommendations are accompanied by positive/negative returns at the time of their announcement.

5.4. Analysts Herding Behavior

We take up the critical question of whether an analyst i tends to follow an analyst j more or less consistently. We randomly draw 200 analysts and plot the coefficients of the dummy variable "buy" at the recommendation revision day in Figure 5. The figures are inferred from (7) for each analyst when each other analyst is included and handled by the α -stable distribution prior and the horseshoe prior¹². Although the mean of the non-zero elements is close to zero, the rest of the descriptive statistics of the non-zero elements reported in each figure imply that some coefficients are very strong in absolute values. The positive coefficients can be explained as the group of analysts who systematically issue similar recommendations. On the contrary, the negative coefficients suggest that some analysts are systematically "contrarian", either because some of them have private information that leads them to issue contrarian recommendations (Bernhardt et al. [5]), or because they want to generate publicity by making extreme recommendations (Ashiya [2]). In other words, a less-talented analyst may mimic the recommendation of a well-

¹²In the Online Appendix, Figures 4-7 gather the results obtained from all the estimation methodologies discussed, presenting the relationships of 200 randomly chosen analysts when they issue "buy" and "sell" recommendations, for $h = 0, 1, 2$.

known analyst or issue a contrarian recommendation, betting on an unanticipated event that could change the firm’s prospects, consequently elevating him to the top of the analysts.

Additionally, we examine if analyst- i follows a list, L_i , of other analysts controlling for NC , adding to the literature of empirical identification of herding. Accordingly, we distinguish between *conditional* and *unconditional* herding. Conditional (unconditional) herding is inferred by the relationships among the analysts’ recommendations coefficients obtained from the estimation of the abnormal returns when we account (do not account) for the effect of NC in the estimation of (7). Assuming that $h = 0$ and the α -stable distribution prior, we plot only the statistically significant coefficients of $I(R_t = 1)$ in Figure 6¹³. Hence, we investigate the structure of “buy” recommendation coefficients and how this is modified when the information conveyed by NC is included or not. In Figure 6(a), the relationships are very sparse conditionally on NC , implying that herding tendencies are very weak, leaving very few significant signals to generate market price reactions. To have a clear view, we provide details for the statistically significant coefficients in Table 5 for the whole sample, including all the MCMC draws. It appears that the α -stable distribution prior is the one that shrinks most of the less important analysts’ effects to zero, as only the 3.5% of the coefficients are not zero. In addition, the percentages of meaningful analysts’ recommendations equal to 12.3%, 7.3%, and 13.1%, when the significance level is $< 1\%$, $1\% - 5\%$, $5\% - 10\%$, respectively. The fractions are quite small, implying that not all analysts are influential, generating abnormal returns. A similar result is found by Loh and Stulz [30], who conclude that only 12% of the recommendation changes are influential.

¹³For $h = 1$ and 2 days holding period, see Online Appendix, Figure 8. For the other recommendations, the same exercise is done.

Table 5: Statistically significant coefficients obtained at conditional herding

	α -stable	Horseshoe	LASSO	ENET
% of non-zero coefficients ($\ A_{norm}\ $)	3.5%	17.2%	15.3%	14.8%
of which “significant” at $< 1\%$	12.3%	21.2%	27.2%	24.3%
of which “significant” at 1%–5%	7.3%	23.35%	32.5%	35.8%
of which “significant” at 5%–10%	13.1%	17.3%	37.8%	40.2%

Note: The table reports the % of non zero coefficients and accordingly, the statistically significant coefficients at 1%, 5% and 10% obtained on the estimation of (7) using different estimation approaches. Numbers represent averages across all MCMC draws.

Figure 6(b) shows that under unconditional herding, the 11% of analysts are influential, when the issue a “buy” recommendation, having in its turn investment value. The relationships among the analysts remain consistent and strong, and analysts tend to follow each other in the absence of consensus. This suggests that unconditional herding is based more on the strength of prior public information and is reinforced by the actions of the market leader, which increases with the level of correlation across informative signals. Based on Graham [21] results, conditional or unconditional herding exists because some analysts have either a greater incentive to hide in the consensus herding¹⁴, or a conservatism in backing the “consensus” of the high reputation analysts to protect their current status.

6. Bayes Factor and Out-of-Sample Behavior

Next, in Figure 7(a)-(c) we present the densities of Bayes factors (BF) in favor of the α -stable model against the LASSO, the elastic net, and the horseshoe prior, respectively, when $B=\{5, 10, 100\}$ analysts are randomly omitted from the data. To obtain these sample distributions, we randomly omit all B analysts’ observations and re-estimate the models. We perform this exercise 500 times. The BF is defined as $BF_{12} = \frac{p(y|\mathcal{M}_1)}{p(y|\mathcal{M}_2)}$, where $p(y|\mathcal{M}_1)$ is the likelihood of the data (when parameter uncertainty has been taken into account) under model \mathcal{M}_1 , and $p(y|\mathcal{M}_2)$ is the probability to observe the data under model \mathcal{M}_2 . BF in favor of the α -stable distribution prior are overwhelming.

¹⁴For example, Clement and Tse [13] present results indicating that less experienced analysts are more likely to herd.

Another interesting issue is how the different models perform in terms of out-of-sample behavior. We take up the problem of model comparison using predictive Bayes factors (Gneiting and Raftery [20]). Given a data set, we split it into an in-sample set for estimation (denoted y_0) and an out-of-sample set (denoted y_1 with $y = y_1 \cup y_0$) which is used for computing the predictive marginal likelihood

$$PML = p(y_1|y_0) = \int p(\theta, y_1|y_0)d\theta = \int p(y_1|\theta, y_0)p(\theta|y_0)d\theta, \quad (11)$$

where θ are the model's parameters. The term $p(\theta|y_0)$ is the posterior of the parameters conditional on the in-sample observations and the multivariate integral in (11) can be accurately approximated as follows. Suppose $\{\theta^{(s)}, s = 1, 2, \dots, S\}$ is a sample that converges to the distribution whose density is $p(\theta|y_0)$. In turn, the integral is approximated using

$$PML(y_1) = S^{-1} \sum_{s=1}^S p(y_1|\theta^{(s)}, y_0). \quad (12)$$

With independent observations, (12) reduces to $PML(y_1) = S^{-1} \sum_{s=1}^S p(y_1|\theta^{(s)})$, which is fairly easy to compute. We denote by PML_0 the predictive marginal likelihood of our new model using the α -stable distribution prior, and PML_j the predictive marginal likelihood using either LASSO, or elastic net, or the horseshoe prior. The predictive Bayes factor (PBF) is defined as

$$PBF_{0:j} = \frac{PML_0(y_1)}{PML_j(y_1)}.$$

We leave aside the 100 last observations from the sample, and we perform MCMC again by introducing these observations in blocks of five observations at a time. Figure 7(d) depicts the findings. In the initial period, all the PBFs are normalized to unity. As we see, the marginal likelihood values and Bayes factors differ widely between LASSO, elastic net, and the horseshoe prior. This is understandable given that different shrinkage methods (LASSO, elastic net) and priors (α -stable, horseshoe) make different assumptions about sparsity and, thus, fit and parsimony.

mony. Clearly, we do not expect different priors to yield the same results, and, especially, the predictions in terms of abnormal returns can be quite different, as we see in our application.

7. Conclusions

This paper introduces a new modelling approach to examine the existence of influential analysts that generate abnormal returns when they issue new recommendations. Given the scarcity of analyst recommendation data, we propose estimating a model for abnormal returns using the α -stable distribution as prior, in conjunction with a time-varying Markov switching model for the analysts' recommendations.

We find that very few analysts' recommendations generate abnormal returns, yet profitable investment strategies based on analysts' recommendations exist. Additionally, we examine whether analysts herd by distinguishing between conditional and unconditional herding. Conditional (unconditional) herding is inferred by the relationships among the analysts' recommendations coefficients obtained from the estimation of the abnormal returns model when we account (do not account) for the effect of the deviation of the analyst's new recommendation from the prevailing consensus. We show that conditional herding is not pervasive given the sparse structure of the coefficient estimates. On the contrary, the structure is very dense when unconditional herding is assumed. Finally, we provide evidence that the estimation using the α -stable distribution performs better than LASSO, elastic net, and the horseshoe prior, both in-sample and out-of-sample.

Future research will focus on building a model combining the information inferred from analysts and brokerages and on identifying the factors differentiating the results obtained from the two datasets.

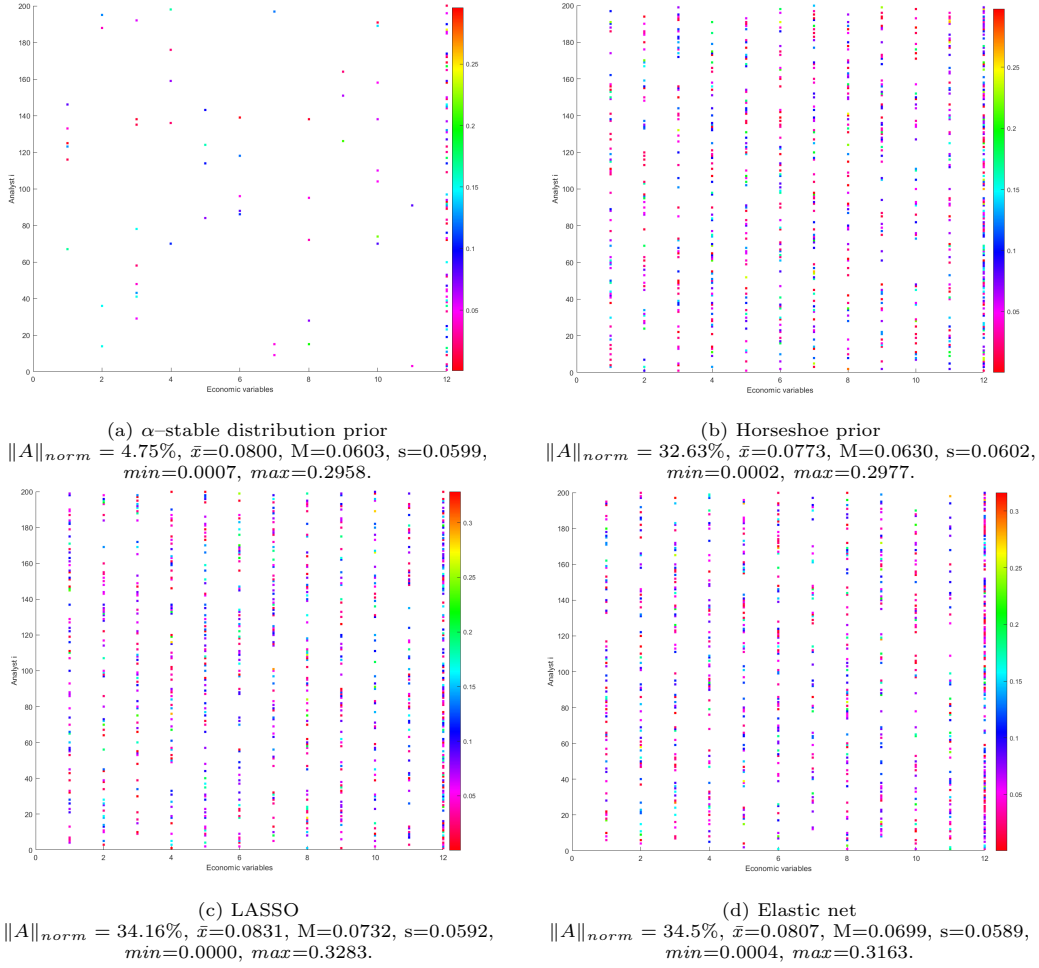


Figure 1: Posterior mean estimates of the coefficients of the market conditions variables¹⁵. *Notes:* Results are based on that $h = 0$. The numbers at the x-axis correspond to the following variables: 1. VIX Index, 2. Price to earnings (P/E) of S&P 500 Index, 3. Growth rate of the trailing 12-month sums of earnings of S&P 500 Index, 4. DEBT/EBITDA ratio of the S&P 500 Index, 5. Net equity expansion on the NYSE Index, 6. Yield to maturity for the three-month Treasury bill, 7. 2/10 U.S. Government bond yield spread, 8. Moody’s BAA and AAA-rated corporate bond yields spread, 9. U.S. inflation, 10. U.S. ISM Manufacturing Purchasing Managers Index, 11. Growth rate of the U.S. Conference Board Leading Economic Indicators Index, 12. U.S. business cycle phase. The y-axis presents 200 randomly chosen analysts.

¹⁵To compare the sparsity among the models, we report the normalized $\|\ell\|_0$ -norm defined by $\|A\|_{norm} := \#\{(i, j), A_{i,j} \neq 0\} / (n * m)$, where $A \in \mathbb{R}^{n * m}$. Also, we report the mean (\bar{x}), the median (M) and the standard deviation (s) of the non-zero elements.

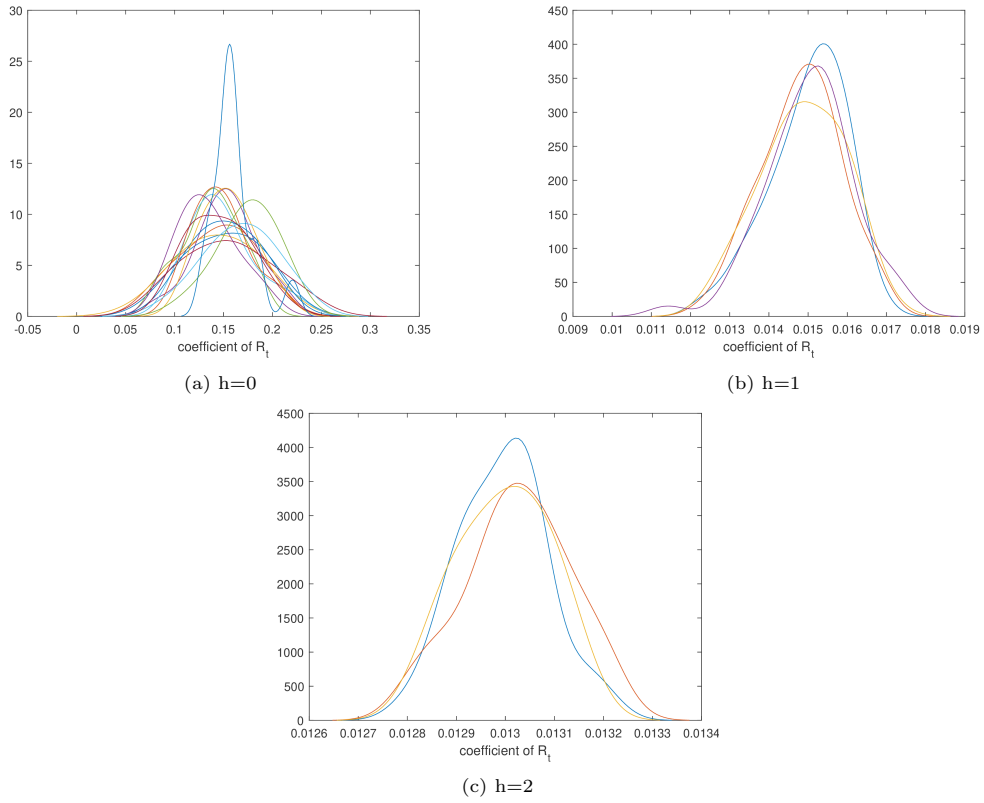


Figure 2: Marginal posterior densities of coefficients of $I(R_t = 1)$. *Notes:* The α -stable distribution prior is used. Only the not zero mass marginal posterior densities are plotted.

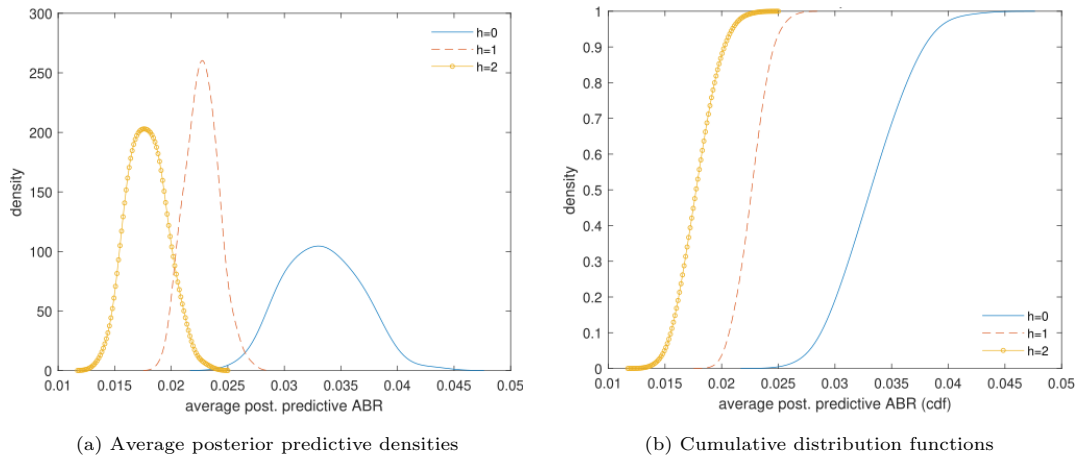


Figure 3: Average posterior predictive densities and cumulative distribution functions of abnormal returns. *Notes:* The densities are obtained using the α -stable distribution prior, for $h = 0, 1, 2$ days after the recommendation revisions.

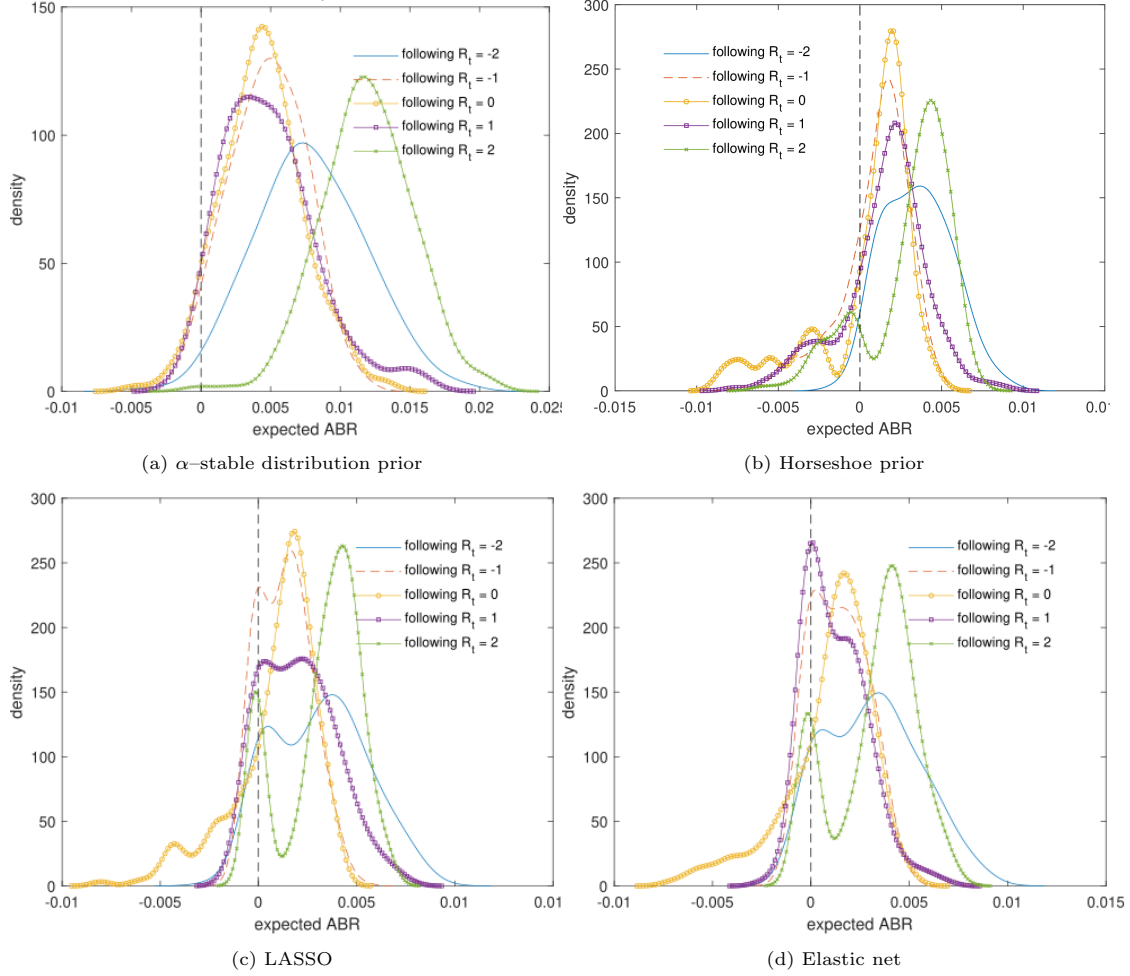


Figure 4: Posterior densities of the expected abnormal returns of the investment strategies. *Note:* It is assumed that $h = 1$.

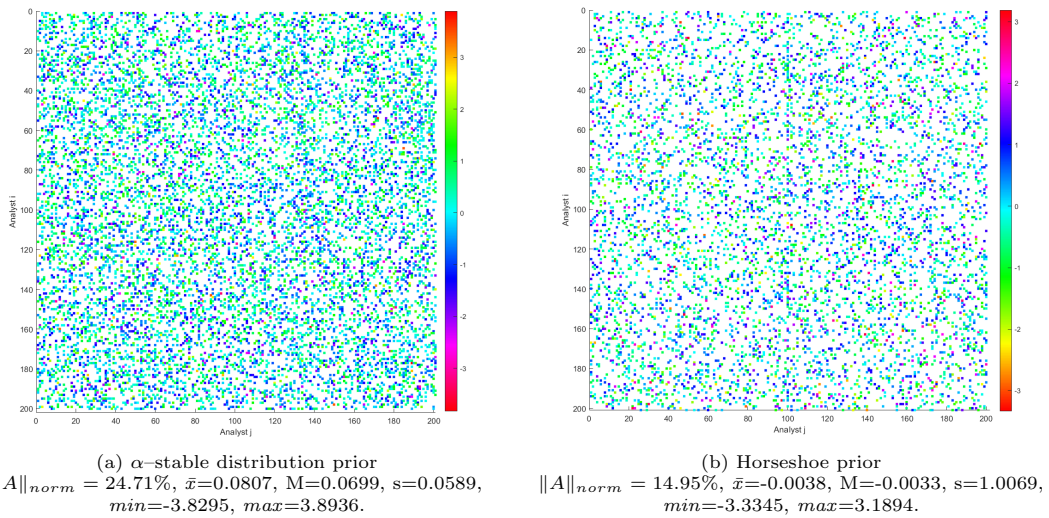


Figure 5: Relations between the coefficients of the variable $I(R_t = 1)$ at the recommendation revision day ($h = 0$)¹⁶.

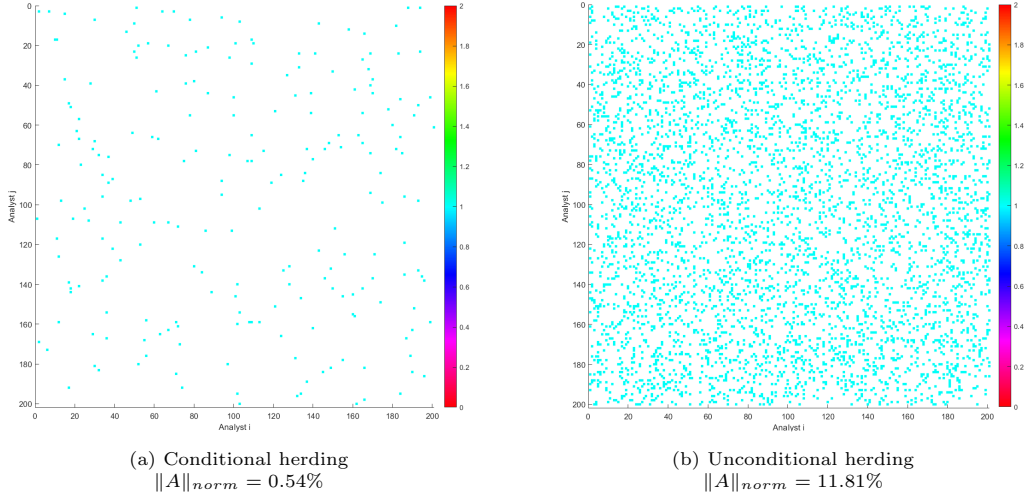


Figure 6: Consistency in following given analysts at the recommendation revision day ($h = 0$). *Notes:* In both cases, only the statistically significant coefficients of $I(R_t = 1)$ assuming the α -stable distribution prior are plotted¹⁴.

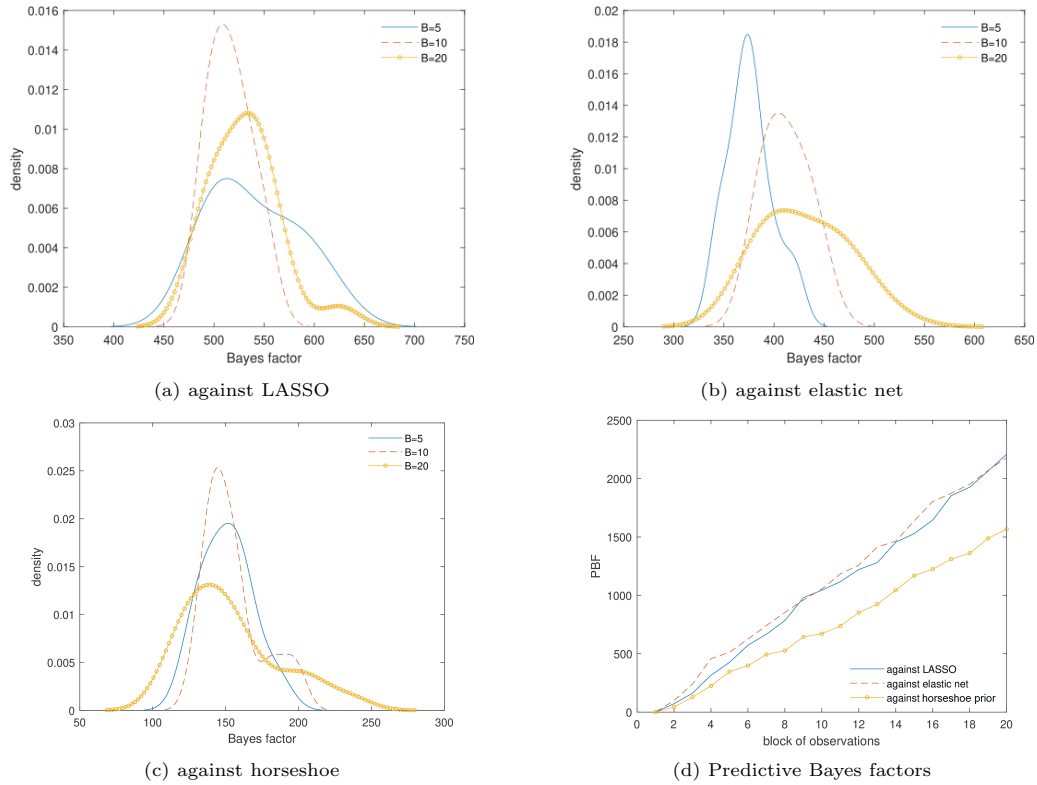


Figure 7: Bayes factor and predictive Bayes factor for model comparison.

¹⁶To compare the sparsity among the models, we report the normalized $\|\ell\|_0$ -norm defined by $\|A\|_{norm} := \#\{(i, j), A_{i,j} \neq 0\} / (n * m)$, where $A \in \mathbb{R}^{n * m}$. Also, we report the mean (\bar{x}), the median (M) and the standard deviation (s) of the non-zero elements.

Acknowledgements

We are grateful to Valerio Poti for his insightful and constructive suggestions that improved the paper in various ways. Also, we thank members of the COST (European Cooperation in Science and Technology) Action 19130 for their comments.

Funding

This paper has received funding from the European Union’s Horizon 2020 research and innovation program “FIN-TECH: A Financial supervision and Technology compliance training programme” under the grant agreement No 825215 (Topic: ICT-35-2018, Type of action: CSA).

References

- [1] Altinkılıç, Oya, Vadim S. Balashov, and Robert S. Hansen (2013) ‘Are analysts’ forecasts informative to the general public?’ *Management Science* 59(11), 2550–2565.
- [2] Ashiya, Masahiro (2009) ‘Strategic bias and professional affiliations of macroeconomic forecasters.’ *Journal of Forecasting* 28(2), 120–30.
- [3] Barber, Brad, Reuven Lehavy, Maureen McNichols, and Brett Trueman (2001) ‘Can investors profit from the prophets? security analyst recommendations and stock returns.’ *Journal of Finance* 56(2), 531–563.
- [4] Bernhardt, Dan, Chi Wan, and Zhijie Xiao (2016) ‘The reluctant analyst.’ *Journal of Accounting and Economics* 54(4), 987–1040.
- [5] Bernhardt, Dan, Murillo Campello, and Edward Kutsoati (2006) ‘Who herds?’ *Journal of Financial Economics* 80(3), 657–675.
- [6] Bhadra, A., J. Datta, N.G. Polson, and B. Willard (2019) ‘Horseshoe regularization for feature subset selection.’ *Sankhya B*.
- [7] Bradshaw, Mark T. (2002) ‘The use of target prices to justify sell-side analysts’ stock recommendations.’ *Accounting Horizons* 16(1), 27–41.

- [8] Brauer, Matthias, and Margarethe Wiersema (2018) ‘Analyzing analyst research: A review of past coverage and recommendations for future research.’ *Journal of Management* 44(1), 218–248.
- [9] Brav, Alon, and Reuven Lehavy (2003) ‘An empirical analysis of analysts’ target prices: Short-term informativeness and long-term dynamics.’ *Journal of Finance* 58(5), 1933–1967.
- [10] Brown, Nerissa C., Kelsey D. Wei, and Russ Wermers (2014) ‘Analyst recommendations, mutual fund herding, and overreaction in stock prices.’ *Management Science* 60(1), 1–20.
- [11] Carvalho, Carlos M., Nicholas G. Polson, and James G. Scott (2010) ‘The horseshoe estimator for sparse signals.’ *Biometrika* 97(2), 465–480.
- [12] Chincó, Alex, Adam D. Clark-Joseph, and Mao Ye (2019) ‘Sparse signals in the cross-section of returns.’ *Journal of Finance* 74(1), 449 – 492.
- [13] Clement, Michael B., and Senyo Y. Tse (2005) ‘Financial analyst characteristics and herding behavior in forecasting.’ *Journal of Finance* 60(1), 307–341.
- [14] Conrad, Jennifer, Bradford Cornwell, Wayne R. Landsman, and Brian R. Rountree (2006) ‘How do analyst recommendations respond to major news?’ *Journal of Financial and Quantitative Analysis* 41(1), 25–49.
- [15] Crane, Alan, and Kevin Crotty (2020) ‘How skilled are security analysts?’ *Journal of Finance* 75(3), 1629–1675.
- [16] Crawford, Steven, Wesley Gray, Bryan R. Johnson, and Richard A. Price III (2017) ‘What motivates buy-side analysts to share recommendations online?’ *Management Science* 64(6), 2574–2589.
- [17] Da, Zhi, Keejae P. Hong, and Sangwoo Lee (2016) ‘What drives target price forecasts and their investment value?’ *Journal of Business Finance & Accounting* 43(3), 487–510.
- [18] Datta, Jyotishka, and Jayanta K. Ghosh (2015) ‘In search of optimal objective priors for model selection and estimation.’ In ‘Current Trends in Bayesian Methodology with Applications’ (New York: Chapman and Hall/CRC) pp. 225–242.
- [19] Gelman, Andrew (2006) ‘Prior distributions for variance parameters in hierarchical models.’

- Bayesian Analysis* 1(3), 515–534.
- [20] Gneiting, Tilmann, and Adrian E Raftery (2007) ‘Strictly proper scoring rules, prediction, and estimation.’ *Journal of the American Statistical Association* 102(477), 359–378.
- [21] Graham, John R. (1999) ‘Herding among investment newsletters: Theory and evidence.’ *Journal of Finance* 54(1), 237–268.
- [22] Green, T. Clifton (2006) ‘The value of client access to analyst recommendations.’ *Journal of Financial and Quantitative Analysis* 41(1), 1–25.
- [23] Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2001) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer Science Business Media, Berlin.)
- [24] Hribar, Paul, and John McInnis (2012) ‘Investor sentiment and analysts’ earnings forecast errors.’ *Management Science* 58(2), 293–307.
- [25] Huang, Joshua, G. Mujtaba Mian, and Srinivasan Sankaraguruswamy (2009) ‘The value of combining the information content of analyst recommendations and target prices.’ *Journal of Financial Markets* 12(4), 754–777.
- [26] Huang, Shiyang, Xin Liu, and Chengxi Yin (2019) ‘Investor target prices.’ *Journal of Empirical Finance* 54, 39–57
- [27] Jegadeesh, Narasimhan, and Woojin Kim (2010) ‘Do analysts herd? an analysis of recommendations and market reactions.’ *Review of Financial Studies* 23(2), 901–937.
- [28] Kim, Yongtae, and Minsup Song (2014) ‘Management earnings forecasts and value of analyst forecast revisions.’ *Management Science* 61(7), 1663–1683.
- [29] Ljungqvist, Alexander, Christopher J. Malloy, and Felicia Marston (2009) ‘Rewriting history.’ *Journal of Finance* 64(4), 1935–1960.
- [30] Loh, Roger K., and René Stulz (2010) ‘When are analyst recommendation changes influential?’ *Review of Financial Studies* 24(2), 593–627.
- [31] Peng, Lin, and Wei Xiong (2006) ‘Investor attention, overconfidence, and category learning.’ *Journal of Financial Economics* 80(3), 563–602.

- [32] Polson, Nicholas G., and James G. Scott (2011) ‘Shrink globally, act locally: Sparse bayesian regularization and prediction.’ *Bayesian Statistics* 9, 501–538.
- [33] — (2012) ‘Local shrinkage rules, lévy processes and regularized regression.’ *Journal of the Royal Statistical Society. Series B* 74(2), 287–311.
- [34] Tibshirani, R. (1996) ‘Regression shrinkage and selection via the lasso.’ *Journal of the Royal Statistical Society. Series B* 58(1), 267–288.
- [35] van der Pas, S.L., B. Szabó, and A. van der Vaart (2016a) ‘How many needles in the haystack? adaptive inference and uncertainty quantification for the horseshoe.’ *arXiv:1607.01892*.
- [36] — (2017) ‘Adaptive posterior contraction rates for the horseshoe.’ *Electron. J. Stat.* (11), 3196–3225.
- [37] van der Pas, S.L., B.J.K. Kleijn, and A. W. van der Vaart (2014) ‘The horseshoe estimator: Posterior concentration around nearly black vectors.’ *Electron. J. Stat.* (8), 2585–2618.
- [38] van der Pas, S.L., J.-B. Salomond, and J. Schmidthieber (2016b) ‘Conditions for posterior contraction in the sparse normal means problem.’ *Electron. J. Stat.* (10), 976–1000.
- [39] Welch, Ivo (2000) ‘Herding among security analysts.’ *Journal of Financial Economics* 58(3), 369–396.
- [40] Womack, Kent L. (1996) ‘Do brokerage analysts’ recommendations have investment value?’ *Journal of Finance* 51(1), 137–167.
- [41] Zellner, Arnold (1971) *An introduction to Bayesian inference in econometrics* (New York: Wiley)
- [42] Zou, H., and T. Hastie (2005) ‘Regularization and variable selection via the elastic net.’ *Journal of the Royal Statistical Society. Series B* 67(2), 301–320.