

# Forecasting dynamic return distributions based on ordered binary choice\*

Stanislav ANATOLYEV<sup>†</sup>      Jozef BARUNÍK<sup>‡</sup>

## Abstract

We present a simple approach to forecasting conditional probability distributions of asset returns. We work with a parsimonious specification of ordered binary choice regression that imposes a connection on sign predictability across different quantiles. The model forecasts the future conditional probability distributions of returns quite precisely when using a past indicator and past volatility proxy as predictors. Direct benefits of the model are revealed in an empirical application to the 29 most liquid U.S. stocks. The forecast probability distribution is translated to significant economic gains in a simple trading strategy. Our approach can also be useful in many other applications where conditional distribution forecasts are desired.

**JEL Classification:** C22, C58, G17

**Keywords:** asset returns, predictive distribution, conditional probability, probability forecasting, ordered binary choice

---

\*We are grateful to the editor Dick van Dijk and two anonymous referees for useful comments and suggestions, which have greatly improved the article.

<sup>†</sup>CERGE-EI, Politických vězňů 7, 11121, Prague, Czech Republic; New Economic School, 45 Skolkovskoe Shosse, Moscow, 121353, Russia.

<sup>‡</sup>Institute of Economic Studies, Charles University, Opletalova 26, 11000, Prague, Czech Republic; Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodarenskou Vezi 4, 18200, Prague, Czech Republic.

# 1 Introduction

*“Those who have knowledge, don’t predict. Those who predict, don’t have knowledge.”*

- Lao Tzu, (c. 604–531 B.C.)

Several decades of research provides overwhelming evidence about the predictability of the first two moments of stock return distributions. Expected values of stock returns are predictable to some extent using economic variables (Keim and Stambaugh, 1986; Fama and French, 1989; Ang and Bekaert, 2006; Viceira, 2012), while the conditional second moment can be well characterized by simple volatility models, or even measured from high frequency data (Bollerslev, 1986; Andersen et al., 2003). While volatility forecasting quickly became central to the financial econometrics literature due to its importance for risk measurement and management, research focusing on the entire return distribution still occupies a small fraction of the literature.<sup>1</sup>

One of the main reasons why researchers may not be focused on characterizing the entire return distribution is a prevailing practice of convenient mean-variance analysis that is still central to modern asset pricing theories. Unfortunately, investor choices guided using the first two moments are restricted by binding assumptions, such as the multivariate normality of stock returns or a quadratic utility function. More importantly, an investor is restricted to have classical preferences based on von Neumann-Morgenstern expected utility. In contrast to this, Rostek (2010) recently developed a notion of quantile maximization and quantile utility preferences. This important shift in decision-theoretic foundations provokes us to depart from the limited mean-variance thinking and work with entire distributions.

Specification and estimation of an entire conditional distribution of future price changes is useful for a number of important financial decisions. Prime examples include portfolio selection when returns are non-Gaussian, (tail) risk measurement and management, market timing strategies with precise entries and exits reflecting information in tails. Despite its importance, forecasting the conditional distribution of future returns has so far attracted little attention in contrast to point forecasts and its aforementioned uncertainty. In this article, we present a simple approach to forecasting a conditional distribution of stock returns using a parameterized ordered binary choice regression. Although focusing on financial returns, we note that our approach may be useful to many other applications where the conditional distribution forecasts are of interest.

The majority of studies focusing on prediction of conditional return distributions characterize the cumulative conditional distribution by a collection of conditional quantiles (Engle

---

<sup>1</sup>Few studies focused on directional forecasts, or threshold exceedances (Christoffersen and Diebold, 2006; Chung and Hong, 2007; Nyberg, 2011).

and Manganelli, 2004; Cenesizoglu and Timmermann, 2008; Žikeš and Baruník, 2016; Pedersen, 2015). In contrast, in a notable contribution, Foresi and Peracchi (1995) focus on a collection of conditional probabilities and describe the cumulative distribution function of excess returns using a set of separate logistic regressions. To be able to approximate the distribution function, Foresi and Peracchi (1995) estimate a sequence of conditional binary choice models over a grid of values corresponding to different points of the distribution. Peracchi (2002) argues that the conditional distributions approach has numerous advantages over the conditional quantile approach, and Leorato and Peracchi (2015) continue their comparison further. The approach has also been considered by Fortin et al. (2011), Rothe (2012), Chernozhukov et al. (2013), Hothorn et al. (2014), and Taylor and Yu (2016).

In this article, we further develop the ideas set forth by Foresi and Peracchi (1995) and present a related simple model for forecasting conditional return distributions. The proposed model is based on an ordered binary choice regression, which is able to forecast the entire predictive distribution of stock returns using fewer parameters than the set of separate binary choice regressions. To achieve this substantial reduction in the degree of parameterization, we tie the coefficients of predictors via smooth dependence on corresponding probability levels. Our specification can be motivated in a semiparametric way, as we approximate smooth probability functions by low-order polynomials.

The probability forecasts are conditional on past information contained in returns, as well as their volatility proxy. The main reason for choosing volatility as one of the explanatory variables is that the cross-sectional relation between risk and expected returns, generally measuring a stock’s risk as the covariance between its return and some factor, is well documented in the literature. In the laborious search for proper risk factors, volatility plays a central role in explaining expected stock returns for decades. Although predictions about expected returns are essential for understanding classical asset pricing, little is known about the potential of these factors to precisely identify extreme tail events of the return distribution.

In our illustrative empirical analysis, we estimate conditional distributions of the 29 most liquid U.S. stocks and compare their generated forecasts with those from the buy-and-hold strategy as well as several benchmarks – a collection of separate binary choice models, a fully specified conditional density, and historical simulation. The benefits of our approach translate into significant economic gains in a simple trading strategy that uses conditional probability forecasts.

We provide the package `DistributionalForecasts.jl` in the Julia software for estimating the model introduced in this article. The package is available at <https://github.com/barunik/DistributionalForecasts.jl>.

The article is organized as follows. Section 2 describes the model and emphasizes its differences with the collection of separate binary choice models. Section 3 contains information about the data we use and lays out details of particular specifications. In Section 4, empirical

results are presented. Section 5 concludes. The Appendix contains more technical material and details of some procedures used in the empirical application.

## 2 Model

We consider a strictly stationary series of financial returns  $r_t$ ,  $t = 1, \dots, T$ . Our objective is to describe, as precisely as possible, the conditional cumulative return distribution  $F(r_t | \mathcal{I}_{t-1})$ , where  $\mathcal{I}_{t-1}$  includes the history of  $r_t$  as well as, possibly, past values of other observable variables.

Consider a partition of the support of returns by  $p > 1$  fixed cutoffs, or thresholds

$$c_1 < c_2 < \dots < c_p,$$

and define  $c_0 = -\infty$  and  $c_{p+1} = +\infty$  for convenience. The higher  $p$ , the more precise the description of the conditional distribution will be (see a full discussion later in the section). The partition  $\{c_j\}_{j=0}^{p+1}$  is arbitrary subject to the ordering restrictions. One intuitive partition corresponds to empirical quantiles of returns: each  $c_j$  is an empirical  $\alpha_j$ -quantile of returns,  $j = 1, \dots, p$ , where  $0 < \alpha_1 < \alpha_2 < \dots < \alpha_p < 1$  are  $p$  probability levels; a reasonable grid for the probability levels is a regularly spaced unit interval  $[0, 1]$ . Alternatively, and perhaps more judiciously, the partition  $\{c_j\}_{j=1}^p$  and thresholds  $\{\alpha_j\}_{j=1}^p$  can be tied to some volatility measure to reflect the time-varying spread of returns due to the changing shape of the conditional distribution. Thus, in general the elements of the partition are time-varying and implicitly indexed by  $t$ .

Let  $\Lambda : u \mapsto [0, 1]$  be a (monotonically increasing) link function. Both unordered and ordered binary choice models are represented by a collection of conditional probabilities

$$\Pr\{r_t \leq c_j | \mathcal{I}_{t-1}\} = \Lambda(\theta_{t,j}), \quad j = 1, \dots, p, \quad (1)$$

for some specification of the driving processes  $\theta_{t,j}$ ,  $j = 1, \dots, p$ . For convenience, define  $\Lambda(\theta_{t,0}) = 0$  and  $\Lambda(\theta_{t,p+1}) = 1$ .

Let  $x_{t-1,j}$ ,  $j = 1, \dots, p$  be a vector of predictors for  $\mathbb{I}_{\{r_t \leq c_j\}}$  that may depend on  $j$  via the dependence of some of them on  $c_j$ . For instance, one of predictors may be  $\mathbb{I}_{\{r_{t-1} \leq c_j\}}$ , the past indicator (dependent on  $j$ ), while another may be  $r_{t-1}$ , the past return (independent of  $j$ ), yet another may represent some volatility measure (also non-specific to  $j$ ). Suppose for simplicity that the number of predictors in  $x_{t-1,j}$  is the same for all  $j$  and equals  $k$ .

In the unordered model, the specification for the underlying process  $\theta_{t,j}$  is

$$\theta_{t,j} = \delta_{0,j} + x'_{t-1,j} \delta_j. \quad (2)$$

There are no cross-quantile restrictions, and each binary choice problem is parameterized separately. This results in a flexible but highly parameterized specification. In the proposed

ordered model, we place cross-quantile restrictions on the parameters. In particular, the coefficients of predictors are tied via smooth dependence on the probability levels; this leads to a substantial decrease in the degree of parameterization.

In the unordered model, no monotonicity, in the language of Foresi and Peracchi (1995), holds in general. That is,  $\Pr\{r_t \leq c_{j-1} | \mathcal{I}_{t-1}\}$  may exceed  $\Pr\{r_t \leq c_j | \mathcal{I}_{t-1}\}$  with positive probability even though  $c_{j-1} < c_j$ . In the proposed ordered model, the monotonicity property in-sample is imposed automatically by the specification of the ordered binary choice likelihood function. This may require artificial adjustments of the conditional distribution values at some thresholds. Out-of-sample, the monotonicity is also not guaranteed to hold, but similar artificial measures can be applied. One simple way is to shift that value of the conditional distribution that violates monotonicity at a particular threshold to its value at the previous threshold plus an additional small figure. An alternative way to ensure both in-sample and out-of-sample predictability is via rearrangement (Chernozhukov et al., 2009). Given the generally low predictability of conditional probabilities for returns (and hence their low variability compared to their mean), a share of observations that need such adjustments is expected to be low (see below for empirical evidence); therefore we give a preference to the former, simpler method.

The specification for the underlying process  $\theta_{t,j}$  is

$$\theta_{t,j} = \delta_{0,j} + x'_{t-1,j} \delta(\alpha_j), \quad (3)$$

where  $\delta(\alpha_j)$  are coefficients that are functions of the probability level  $\alpha_j$ . Each probability-dependent slope coefficient vector is specified as follows:  $\delta(\alpha_j) = (\delta_1(\alpha_j), \dots, \delta_k(\alpha_j))'$ , where for each  $\ell = 1, \dots, k$ ,

$$\delta_\ell(\alpha_j) = \kappa_{0,\ell} + \sum_{i=1}^{q_\ell} 2^i (\alpha_j - 0.5)^i \cdot \kappa_{i,\ell}, \quad (4)$$

and  $q_\ell \leq p - 1$ . Note that each intercept  $\delta_{0,j}$  is  $j$ -specific and represents an ‘individual effect’ for a particular probability level, while the slopes  $\delta$ ’s do not have index  $j$ , i.e. they depend on  $j$  only via dependence on  $\alpha_j$ ’s. The motivation behind such specification is semiparametric: any smooth function on  $[0, 1]$  can be approximated to a desired degree of precision by the system of basis polynomials  $\{\alpha_j - 0.5, (\alpha_j - 0.5)^2, \dots, (\alpha_j - 0.5)^q\}$  by making  $q$  big enough. Because all  $\alpha_j \in (0, 1)$ , the polynomial form behaves nicely even for a large  $q$ ; the basis polynomials are uniformly bounded on  $[0, 1]$ . The additional weights  $2^i$  are introduced to line up the coefficients  $\kappa_i$  on a more comparable level.

Let us compare the degrees of parameterization of the unordered and ordered binary choice models. Denote

$$q = \sum_{\ell=1}^k q_\ell.$$

In the unordered model, the total number of parameters is

$$K_{UO} = (1 + k)p$$

(namely, one intercept  $\delta_{0,j}$  and  $k$  slopes  $\delta_j$  in each of  $p$  equations for  $\theta_j$ ), while in the ordered model, the total number of parameters is

$$K_O = p + k + q$$

(namely,  $p$  intercepts  $\delta_{0,j}$  and  $k$  slopes  $\delta(\alpha_j)$ , each parameterized via  $1 + q_\ell$  parameters). The difference

$$K_{UO} - K_O = k(p - 1) - q$$

is larger the larger  $p$  is, the fineness of the partition by thresholds. The resulting difference is also positively related to the number of predictors used.<sup>2</sup>

In the unordered model, the composite loglikelihood corresponding to observation  $t$  is

$$\ell_t^{UO} = \sum_{j=1}^{p+1} \mathbb{I}_{\{r_t \leq c_j\}} \ln(\Lambda(\theta_{t,j})), \quad (5)$$

and the total composite likelihood  $\sum_{t=1}^T \ell_t^{UO}$  can be split into  $p$  independent likelihoods  $\sum_{t=1}^T \ell_t^{(j)}$ , where

$$\ell_t^{(j)} = \mathbb{I}_{\{r_t \leq c_j\}} \ln(\Lambda(\theta_{t,j})), \quad (6)$$

to be maximized over the parameter vector  $(\delta_{0,j}, \delta_j)'$ . In the ordered model, the loglikelihood corresponding to observation  $t$  is

$$\ell_t^O = \sum_{j=1}^{p+1} \mathbb{I}_{\{c_{j-1} < r_t \leq c_j\}} \ln(\Delta_j \Lambda_t), \quad (7)$$

where  $\Delta_1 \Lambda_t = \Lambda(\theta_{t,1})$ ,  $\Delta_j \Lambda_t = \Lambda(\theta_{t,j}) - \Lambda(\theta_{t,j-1})$  for  $j = 2, \dots, p$ , and  $\Delta_{p+1} \Lambda_t = 1 - \Lambda(\theta_{t,p})$ . The total likelihood  $\sum_{t=1}^T \ell_t^O$  is to be maximized over the parameter vectors  $(\delta_{0,1}, \dots, \delta_{0,p})'$  and  $(\kappa_{0,1}, \dots, \kappa_{0,k}, \kappa_{1,1}, \dots, \kappa_{q_k, k})'$ . Under mild suitable conditions, the estimates of the parameter vector are expected to be consistent for their pseudotrue values and asymptotically normal around them, with a familiar sandwich form of the asymptotic variance.

Of course, each difference  $\Delta_j \Lambda_t$  needs to be positive. This monotonicity property, though not guaranteed, is easier to enforce in maximization of the joint ordered likelihood  $\sum_{t=1}^T \ell_t^O$  than in separate maximizations of independent unordered likelihoods, as the common parameters will automatically adjust to these monotonicity restrictions. However, if the degrees of freedom are insufficient to ensure this for all predictor values in the sample, to prevent

---

<sup>2</sup>In our empirical illustration,  $p = 37$ ,  $k = 2$  and  $q_1 = 2$ ,  $q_2 = 3$ . Hence,  $K_{UO} = 111$  while  $K_O = 44$ , and the difference is  $K_{UO} - K_O = 67$ .

(rare) realizations of negative differences for particular  $t$ , monotonicity can be enforced by imposing constraints  $\Delta_j \Lambda_t \geq \varepsilon$  for all  $j$  and  $t$ , where  $\varepsilon$  is some small number.<sup>3</sup>

Computationally, it is convenient to maximize the total likelihood in a number of steps. The reason is that an arbitrary initial parameter vector is likely to result in incomputable likelihood because of numerous violations of the monotonicity property. The idea is to first determine approximate values of individual intercepts and slopes, subject to their monotonicity, then relaxing the restrictions on the slopes using the evaluated values as starting points for corresponding parts of the parameter vector. Towards this end, we propose and further use the following algorithm<sup>4</sup>:

- Step 1. Run a series of separate binary choice models (1) and (2) with the specification  $\theta_{t,j} = \delta_{0,j} + x'_{t-1,j} \delta_j$  with  $\delta_j = (\delta_{j,1}, \dots, \delta_{j,k})'$ ,  $j = 1, \dots, p$ , by maximizing the individual likelihoods (6), and call the obtained estimates  $\bar{\delta}_{0,j}$  and  $\bar{\delta}_j$ .
- Step 2. For each  $\ell = 1, \dots, k$ , run a linear regression  $\bar{\delta}_\ell = \kappa_{0,\ell} + \sum_{i=1}^{q_\ell} 2^i (\alpha - 0.5)^i \cdot \kappa_{i,\ell}$ , where  $\delta_\ell = (\delta_{1,\ell}, \dots, \delta_{p,\ell})'$  and  $\alpha = (\alpha_1, \dots, \alpha_p)'$ , and call the obtained estimates  $\bar{\kappa}_{i,\ell}$ ,  $i = 0, 1, \dots, q_\ell$ .
- Step 3. Run the ordered binary choice model (1) and (3) with specification (4) by maximizing the total likelihood (7) using  $\bar{\delta}_{0,j}$  as starting points for  $\delta_{0,j}$ ,  $j = 1, \dots, p$ , and  $\bar{\kappa}_{i,\ell}$  as starting points for  $\kappa_{i,\ell}$ ,  $i = 0, 1, \dots, q_\ell$ ,  $\ell = 1, \dots, k$ .

Having estimated the conditional return distribution evaluated at the threshold values, one can obtain the entire (continuous) conditional distribution by using interpolation schemes that preserve monotonicity of the outcome. Towards this end, we apply the Fritsch–Carlson monotonic cubic interpolation (Fritsch and Carlson, 1980) (see Appendices A.1 and A.2) and use the result for testing the quality of the estimated distribution (see Appendices A.3 and A.4).

The quality of approximation of the constructed conditional distribution in the ordered model is determined by a number of factors, the precision of interpolation being the least important. There is an important tradeoff between the number of thresholds (and hence the precision of the interpolation input), and the degree of parameterization and hence the amount of estimation noise. Yet another factor is the flexibility of specification of the slopes  $\delta$ 's on  $\alpha_j$ 's. It seems reasonable to set the system of thresholds fine enough (as long as one does not come close to the computability limits) to describe the distribution precisely enough

---

<sup>3</sup>In our empirical illustration, we made each difference to be bounded below by  $\varepsilon = 10^{-6}$  by force. Such strategy has resulted in only less than 1% of such interferences among all differences during in-sample estimation and less than 2% during out-of-sample forecasting.

<sup>4</sup>The estimation can be done using the package `DistributionalForecasts.jl` developed by the authors in the Julia software. The package is available at <https://github.com/barunik/DistributionalForecasts.jl>

but not too fine so that there are a sufficient number of observations that fall between each pair of adjacent thresholds. One may also afford higher flexibility to the slope specification for larger sample sizes; however, low numbers are usually pretty adequate in semiparametric setups in practice. One may also employ formal model selection criteria such as the Bayesian information criterion to choose the optimal orders of polynomials in slope specifications.

### 3 Data and Empirical Specification

We study the conditional distribution forecasts of 29 U.S. stocks<sup>5</sup> that are traded on the New York Stock Exchange. These stocks have been chosen according to market capitalization and their liquidity. The sample under investigation spans from August 19, 2004 to December 31, 2015. We consider trades executed during U.S. business hours (9:30–16:00 EST). In order to ensure sufficient liquidity and eliminate possible bias we explicitly exclude weekends and bank holidays (Christmas, New Year’s Day, Thanksgiving Day, Independence Day). In total, our final dataset consists of 2826 trading days, 500 of which are used for in-sample estimation, and the rest 2326 for out-of-sample forecasting using a rolling window scheme with the window size of 500 days. We split the sample to have a much larger out-of-sample portion as we perform an extensive set of tests, robustness checks and inter-model comparisons on it.

Next we give the details of our empirical specification. For both ordered and unordered models, we use the logit link function

$$\Lambda(u) = \frac{\exp(u)}{1 + \exp(u)}$$

resulting in logit specifications. We consider a partition of the return space into 37 equally spaced probability levels<sup>6</sup> from  $\alpha \in (5\%, 95\%)$ , i.e., a grid with step 2.5% resulting in  $p = 37$  quantiles in total.<sup>7</sup> We use a time-varying partition that changes with the rolling window. In every window,  $c_j$  is computed as  $\gamma(\alpha_j)\sqrt{\sigma_t^2}$ , where  $\gamma(\alpha_j)$  is a quantile of the standard

---

<sup>5</sup>Apple Inc. (AAPL), Amazon.com, Inc. (AMZN), Bank of America Corp (BAC), Comcast Corporation (CMCSA), Cisco Systems, Inc. (CSCO), Chevron Corporation (CVX), Citigroup Inc. (C), Walt Disney Co (DIS), General Electric Company (GE), Home Depot Inc. (HD), International Business Machines Corp. (IBM), Intel Corporation (INTC), Johnson & Johnson (JNJ), JPMorgan Chase & Co. (JPM), The Coca-Cola Co (KO), McDonald’s Corporation (MCD), Merck & Co., Inc. (MRK), Microsoft Corporation (MSFT), Oracle Corporation (ORCL), PepsiCo, Inc. (PEP), Pfizer Inc. (PFE), Procter & Gamble Co (PG), QUALCOMM, Inc. (QCOM), Schlumberger Limited. (SLB), AT&T Inc. (T), Verizon Communications Inc. (VZ), Wells Fargo & Co (WFC), Wal-Mart Stores, Inc. (WMT), Exxon Mobil Corporation (XOM).

<sup>6</sup>The use of more extreme values of probability levels such as 1% and 99% leads, with our in-sample window of 500 observations, to large estimation uncertainty in the tails, but ends up successful when the model is estimated on the whole sample.

<sup>7</sup>We have also used partitions with  $p = 19$  and  $p = 73$  equally spaced probability levels; for details, see subsection 4.4.



normal distribution, and  $\sigma_t^2$  is a conditional variance of returns in the corresponding window computed from the RiskMetrics of JPMorgan Chase standards as an exponentially weighted moving average with a decay factor of 0.94.

We choose  $k = 2$  and the predictors to be

$$x_{t-1,j} = \begin{pmatrix} \mathbb{I}_{\{r_{t-1} \leq c_j\}} \\ \ln(1 + |r_{t-1}|) \end{pmatrix}$$

for all  $j = 1, \dots, p$ . The first predictor is a lagged indicator corresponding to the probability level  $\alpha_j$ , which is plausibly supposed to have the highest predictability power among all such indicators. The second predictor is a proxy for a volatility measure, with the absolute return dampened by the logarithmic transformation. Note that the first predictor is specific to a specific quantile, while the second predictor is common for all quantiles. In principle, one could specify all predictors to be the same across the quantiles, or, on the opposite, all predictors may vary with the quantile. In the ordered model, we set, after some experimentation with statistical significance of higher-order polynomials,  $q_1 = 2$  and  $q_2 = 3$ .<sup>8</sup> That is, the polynomial is quadratic in the probability level  $\alpha$  for the past indicator and cubic for the past volatility proxy.

The full specification of the model for  $j = 1, \dots, p$  empirical quantiles is

$$\Pr\{r_t \leq c_j | \mathcal{I}_{t-1}\} = \frac{\exp(\theta_{t,j})}{1 + \exp(\theta_{t,j})},$$

$$\theta_{t,j} = \delta_{0,j} + \delta_1(\alpha_j) \mathbb{I}_{\{r_{t-1} \leq c_j\}} + \delta_2(\alpha_j) \ln(1 + |r_{t-1}|),$$

with the coefficient functions

$$\delta_1(\alpha_j) = \kappa_{0,1} + 2(\alpha_j - 0.5) \cdot \kappa_{1,1} + 2^2(\alpha_j - 0.5)^2 \cdot \kappa_{2,1},$$

$$\delta_2(\alpha_j) = \kappa_{0,2} + 2(\alpha_j - 0.5) \cdot \kappa_{1,2} + 2^2(\alpha_j - 0.5)^2 \cdot \kappa_{2,2} + 2^3(\alpha_j - 0.5)^3 \cdot \kappa_{3,2}.$$

In total, there are  $K_O = 44$  parameters:  $p = 37$  individual intercepts  $\delta_{0,j}$  and  $k = 2$  slopes  $\delta(\alpha_j)$ , one parameterized via  $1 + q_1 = 3$  parameters, the other via  $1 + q_2 = 4$  parameters. This parametrization is parsimonious enough and approximates the distribution quite well, and additional terms do not bring significant improvements. Hence, estimating seven parameters  $\kappa_{i,\ell}$  in addition to individual intercepts is enough to approximate the conditional return distribution.

## 4 Empirical Findings

We now present the results of estimating the conditional distribution function of returns using the ordered binary choice model. We consider forecasts for 29 stocks, so first we

---

<sup>8</sup>We also perform model selection analysis with the Bayesian information criterion in the ordered model; for details, see subsection 4.4.

present individual estimates of three illustrative stocks, namely Intel Corporation (INTC), QUALCOMM, Inc. (QCOM), and Exxon Mobil Corporation (XOM), and then we present the results for all 29 stocks in an aggregate form.

After presenting the parameter estimates, we evaluate the statistical as well as economic significance of the predicted distributions. Further, we compare the performance of the ordered model with popular and challenging benchmarks used in the literature. Namely, we include two candidate models that perform best (Kuester et al., 2006) – an asymmetric generalized autoregressive conditional heteroscedastic model with skewed-t distribution (GARCH henceforth) and GARCH-filtered historical simulation (FHS henceforth). In implementing both alternative methods, we follow Kuester et al. (2006).

## 4.1 Parameter estimates

Table 1 shows estimates of cutoff-specific intercepts  $\delta_{0,j}$  in the ordered logit model for the three illustrative stocks. The intercepts have expected signs – negative for probability levels to the left of 50% and positive to the right of 50% – due to monotonically increasing  $\Lambda(\cdot)$  and low predictability of predictors and exhibit expected monotonic behavior increasing from the left tail to the right tail, thus generating an increasing cumulative distribution function (assuming zero predictors). The intercepts are statistically significant in most cases except for few a cutoff points near the median. The intercept values are quite similar across three stocks, though there is some variability.

The left plot shown in Figure 2 collects the intercept estimates for all 29 stocks and reveals that indeed the intercepts are similar across the stocks. The plot has a shape similar to an inverse logit link function  $\Lambda(u)$  defined earlier, with a stronger effect in the tails. A flexible  $j$ -specific intercept allows for controlling individual quantile effects, and shows that unconditional expectations are an important part of the predicted distribution.

The estimates of the seven coefficients  $\kappa_{i,\ell}$  driving the slopes on the predictors for the three illustrative stocks, are shown in Table 2. The coefficients  $\kappa_{i,1}$  correspond to the lagged indicator predictor  $\mathbb{I}_{\{r_{t-1} \leq c_j\}}$ . All parameter estimates for the lagged indicator are of a small magnitude, and some are highly statistically significant. The left plot in Figure 1 complements these findings with estimates of the lagged indicator coefficients for all 29 stocks shown by the box-and-whisker plots. The estimates document a heterogeneous effect of the lagged indicator on the future probabilities for different stocks. While we document zero coefficients for some stocks, the quadratic term seems to play a big role in others.

The second predictor, the past volatility proxy  $\ln(1 + |r_{t-1}|)$ , carries even more important information about future probabilities. The estimated coefficients  $\kappa_{i,2}$  corresponding to the volatility predictor reveal that the cubic polynomial has many statistically significant coefficients (see Table 2). The right plot of Figure 1 shows estimated coefficients for all 29 stocks as box-whisker plots. We can see that  $\kappa_{1,2}$  and  $\kappa_{3,2}$  are significantly different from zero

for most of the stocks, and so the past volatility proxy contributes strongly to the prediction of return distributions.

Figure 2 plots the functions  $\delta_1(\alpha_j)$ , and  $\delta_2(\alpha_j)$ , in addition to the intercepts. Corresponding to heterogeneous parameters  $\kappa_{i,1}$ , the function  $\delta_1(\alpha_j)$  is also heterogeneous for 29 stocks, exhibiting a mixed effect of the lagged indicator predictor (shown in the middle plot of Figure 2). Overall, we can see that the effect is small. Coefficient function  $\delta_2(\alpha_j)$  implied by the past volatility proxy shows a similar impact.

Figure 3 depicts an interpolated predictive conditional CDF of returns on the interval  $[-1.5, 1.5]$  for an arbitrary stock<sup>9</sup> evaluated at arbitrary 100 out-of-sample periods. One can observe how the cumulative distribution varies over time. There is a certain distribution clustering, i.e. the CDF possesses some persistence, while at some periods the CDF shape stands out from the cluster.

## 4.2 Statistical Fit

To assess the adequacy of the predicted distribution, we run the generalized autocontours specification tests of González-Rivera and Sun (2015). This test verifies whether the collection of out-of-sample generalized residuals (also known as probability integral transform) together with their lags are scattered uniformly inside a hypercube of corresponding dimension; see Appendix A.3 for a detailed description. We use lag  $L = 1$  in the contour-aggregated test with a simple collection of sides  $\alpha = (0.25, 0.50, 0.75)'$  and a larger collection coinciding with our full grid  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_p)'$ . We use side  $\bar{\alpha} = 0.5$  in the lag-aggregated test with  $L = 3$  and  $L = 10$ . Figure 4 shows the distributions of p-values from the González-Rivera and Sun (2015) tests for the ordered logit model for all 29 stocks. All four variations of the test – two contour-aggregated and two lag-aggregated – show the adequacy of the estimated conditional distribution.<sup>10</sup>

We also compare probabilistic forecasts of different models in terms of proper scoring rules (Gneiting and Raftery, 2007), namely the Brier score and the continuous ranked probability score (CRPS); see Appendix A.4 for a detailed description.<sup>11</sup> Figure 5 shows the average score values for the four models (ordered logit, collection of separate logits, GARCH and FHS). All four approaches deliver Brier score that have similar median values though differing a bit in dispersion. The CRPS scores are very similar across all four approaches, though the logit models marginally dominate.

---

<sup>9</sup>The minimal and maximal values of returns in the whole sample for this stock are approximately  $-0.20$  and  $0.20$ .

<sup>10</sup>We did not account for the estimation noise in constructing the test. Because it generally tends to increase asymptotic variances, the p-values would be even higher if the estimation noise was accounted for.

<sup>11</sup>We follow Gneiting and Raftery (2007) and define the CRPS with a minus sign so that its larger values are preferred to smaller values.

### 4.3 Economic Significance

To understand the economic usefulness of the proposed model, we study a simple profit rule for timing the market based on the model forecasts. The idea is to explore information from the entire distribution. To build a trading strategy, we explore the difference between the predicted conditional probability and unconditional probability, which indicates if positive or negative returns are predicted with higher probability. In the spirit of the previous literature (Breen et al., 1989; Pesaran and Timmermann, 1995; Anatolyev and Gospodinov, 2010), we evaluate the model forecasts in terms of profits from a trading strategy for active portfolio allocation between holding a stock and investing in a risk-free asset. The detailed construction of the trading strategy is described in Appendix A.5.

Table 3 summarizes the results of running the trading strategies in terms of a mean and median return, volatility, as well as Sharpe ratio. While the benchmark market strategy earns a 0.806 return with a 0.244 volatility, GARCH and FHS generate similar returns with a lower volatility. When an investor uses predictions about the entire distribution from separate logit models, an improved mean return of 1.088 is obtained with a much lower volatility. Finally, our proposed ordered logit model generates a 1.296 mean return with a volatility similar to that of separate logits. The Sharpe ratio showing a risk-adjusted return, or a average return per unit of volatility, reveals that the ordered logit model generates the highest returns when one takes risk into account. The figures for median values confirm this result.

Figure 6 shows returns, volatilities, and Sharpe ratios for all 29 stocks using box-and-whisker plots. The figure reveals that the naive buy and hold strategy yields low returns with the highest volatility, which are moreover quite heterogeneous for all the considered stocks. GARCH and FHS improve the volatility estimates, hence yielding similar returns with a lower risk. In terms of risk-adjusted Sharpe ratios, the separate logistic regressions yield a similar result, while the ordered logit model makes a marked improvement. One can see positive Sharpe ratios for almost all considered stocks pointing to the highest risk-adjusted returns offered by the ordered model.

Figure 7 looks closely at the cumulative returns and drawdowns of the three illustrative stocks we used earlier. It can be seen that returns from the ordered logit strategy are consistent over time, with the lowest drawdown. This holds even in the case of a XOM that has been growing for the whole period, hence making it difficult to beat the buy and hold strategy.

Finally, we compare the relative performance of all five strategies in Figure 8. The top left plot in the Figure compares the ordered logit with the ‘market,’ or naive buy and hold strategy, while the top right plot compares the ordered logit with separate logits. The plots at the bottom of Figure 8 compare the ordered logit versus GARCH and FHS. We document a consistently better performance for the proposed ordered logit model in comparison with

both the unordered logit as well as benchmark strategies for all 29 stocks.

## 4.4 Sensitivity Analysis

As noted before, we have chosen, after some experimentation, specification with  $p = 37$  thresholds and polynomial orders  $q_1 = 2$  and  $q_2 = 3$ . Here we report on the results with different specifications. As a robustness check, we do an investigation of the impact of partition fineness and polynomial complexity on the performance of the proposed method.

We have estimated and evaluated the model with different choices of partitions, namely  $p = 19$  (corresponding to twice as coarse a partition) and  $p = 73$  (corresponding to twice as fine a partition). On the one hand,  $p = 19$  may seem too coarse for the CDF approximation to be considered good. On the other hand, too high  $p$  may seriously impact the model complexity: while the number of parameters for the basic partition  $p = 37$  equals  $K_O = 44$ , for the finer partition it equals 80, which is obviously too many for our out-of-sample exercise with 500 rolling in-sample observations. In our experiments with  $p = 19$  and  $p = 73$ ,<sup>12</sup> some of the generalized autocontours tests (in particular, the three-side contour-aggregated test) exhibits the tendency to reject the constructed conditional distribution. The results of economic evaluation though are not sensitive to the choice of partition, and the dominance of the proposed method based on the ordered logit over all the other methods still prevails.

The polynomial orders  $(q_1, q_2)$  do not affect the degree of parsimony drastically, on the one hand, and on the other, the dependence of predictability on the probability level is presumably not very sophisticated to require higher-order powers. Hence, one would not expect high sensitivity to the choice of orders. We have run the proposed model with various combinations of polynomial orders  $(q_1, q_2)$  around the running  $(q_1, q_2) = (2, 3)$  combination, and computed the value of the Bayesian information criterion (BIC) for each. The pattern of BIC is presented in Figure 9. While the values of  $q_1$  and  $q_2$  such as 0 and 1 are obviously too small to capture the differences in predictability across the probability levels, there is little sensitivity once the orders reach the selected combination, which is clearly preferred by BIC en masse across the 29 stocks.

## 5 Conclusion

This article investigates the predictability of stock market return distributions. We propose a relatively parsimonious parametrization of an ordered binary choice regression that forecasts well the conditional probability distribution of returns. We subject the proposed model to a number of statistical tests for adequacy and to comparison with alternative methods. In order to see how useful the model is economically, we use distributional predictions in a

---

<sup>12</sup>The detailed results are available upon request from the authors.

simple market timing strategy. Using 29 liquid U.S. stocks, we find significant economic gains when compared to the benchmarks.

Our findings are useful for risk management and measurement or building trading strategies using the entire conditional distribution of returns. The model nevertheless has a much wider potential use in any application that exploits distribution forecasts, including forecasting of interest rates, term structures, and macroeconomic variables.

## References

- Anatolyev, S. and N. Gospodinov (2010). Modeling financial return dynamics via decomposition. *Journal of Business & Economic Statistics* 28(2), 232–245.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica* 71(2), 579–625.
- Ang, A. and G. Bekaert (2006). Stock return predictability: Is it there? *Review of Financial Studies* 20(3), 651–707.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31(3), 307–327.
- Breen, W., L. R. Glosten, and R. Jagannathan (1989). Economic significance of predictable variations in stock index returns. *Journal of Finance* 44(5), 1177–1189.
- Cenesizoglu, T. and A. G. Timmermann (2008). Is the distribution of stock returns predictable? Manuscript, UCSD.
- Chernozhukov, V., I. Fernández-Val, and A. Galichon (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika* 96(3), 559–575.
- Chernozhukov, V., I. Fernández-Val, and B. Melly (2013). Inference on counterfactual distributions. *Econometrica* 81(6), 2205–2268.
- Christoffersen, P. F. and F. X. Diebold (2006). Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science* 52(8), 1273–1287.
- Chung, J. and Y. Hong (2007). Model-free evaluation of directional predictability in foreign exchange markets. *Journal of Applied Econometrics* 22(5), 855–889.
- Engle, R. F. and S. Manganelli (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics* 22(4), 367–381.
- Fama, E. F. and K. R. French (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* 25(1), 23–49.
- Foresi, S. and F. Peracchi (1995). The conditional distribution of excess returns: An empirical analysis. *Journal of American Statistical Association* 90(430), 451–466.
- Fortin, N., T. Lemieux, and S. Firpo (2011). Decomposition methods in economics. *Handbook of Labor Economics* 4, 1–102.

- Fritsch, F. and R. Carlson (1980). Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis* 17(2).
- Gneiting, T. and A. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of American Statistical Association* 102(477), 359–378.
- González-Rivera, G. and Y. Sun (2015). Generalized autocontours: Evaluation of multivariate density models. *International Journal of Forecasting* 31(3), 799–814.
- Hothorn, T., T. Kneib, and P. Bühlmann (2014). Conditional transformation models. *Journal of Royal Statistical Society: Series B (Statistical Methodology)* 76(1), 3–27.
- Judd, K. (1998). *Numerical Methods in Economics*. Cambridge, MA: MIT Press.
- Keim, D. B. and R. F. Stambaugh (1986). Predicting returns in the stock and bond markets. *Journal of Financial Economics* 17(2), 357–390.
- Kuester, K., S. Mittnik, and M. Paolella (2006). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics* 4(1), 53–89.
- Leorato, S. and F. Peracchi (2015). Comparing distribution and quantile regression. Technical report, Einaudi Institute for Economics and Finance.
- Nyberg, H. (2011). Forecasting the direction of the us stock market with dynamic binary probit models. *International Journal of Forecasting* 27(2), 561–578.
- Pedersen, T. Q. (2015). Predictable return distributions. *Journal of Forecasting* 34(2), 114–132.
- Peracchi, F. (2002). On estimating conditional quantiles and distribution functions. *Computational statistics & data analysis* 38(4), 433–447.
- Pesaran, M. H. and A. Timmermann (1995). Predictability of stock returns: Robustness and economic significance. *Journal of Finance* 50(4), 1201–1228.
- Rostek, M. (2010). Quantile maximization in decision theory. *Review of Economic Studies* 77(1), 339–371.
- Rothe, C. (2012). Partial distributional policy effects. *Econometrica* 80(5), 2269–2301.
- Taylor, J. W. and K. Yu (2016). Using auto-regressive logit models to forecast the exceedance probability for financial risk management. *Journal of Royal Statistical Society: Series A (Statistics in Society)* 179(4), 1069–1092.
- Viceira, L. M. (2012). Bond risk, bond return volatility, and the term structure of interest rates. *International Journal of Forecasting* 28(1), 97–117.



Žikeš, F. and J. Baruník (2016). Semi-parametric conditional quantile models for financial returns and realized volatility. *Journal of Financial Econometrics* 14(1), 185–226.

# A Appendix

## A.1 CDF interpolation

The Fritsch–Carlson monotonic cubic interpolation (Fritsch and Carlson, 1980) provides a monotonically increasing CDF with range  $[0, 1]$  when applied to CDF estimates on a finite grid.

Suppose we have CDF  $F(r)$  defined at points  $(r_k, F(r_k))$  for  $k = 1, \dots, K$ , where  $F(r_0) = 0$  and  $F(r_K) = 1$ . We presume that  $r_k < r_{k+1}$  and  $F(r_k) < F(r_{k+1})$  for all  $k = 0, \dots, K - 1$ , which is warranted by the continuity of returns and construction of the estimated distribution. First, we compute the slopes of the secant lines as  $\Delta_k = (F(r_{k+1}) - F(r_k)) / (r_{k+1} - r_k)$  for  $k = 1, \dots, K - 1$ , and then the tangents at every data point as  $m_1 = \Delta_1$ ,  $m_k = \frac{1}{2}(\Delta_{k-1} + \Delta_k)$  for  $k = 2, \dots, K - 1$ , and  $m_K = \Delta_{K-1}$ . Let  $\alpha_k = m_k / \Delta_k$  and  $\beta_k = m_{k+1} / \Delta_k$  for  $k = 1, \dots, K - 1$ . If  $\alpha_k^2 + \beta_k^2 > 9$  for some  $k = 1, \dots, K - 1$ , then we set  $m_k = \tau_k \alpha_k \Delta_k$  and  $m_{k+1} = \tau_k \beta_k \Delta_k$ , with  $\tau_k = 3(\alpha_k^2 + \beta_k^2)^{-1/2}$ . Finally, the cubic Hermite spline is applied: for any  $r \in [r_k, r_{k+1}]$  for some  $k = 0, \dots, K - 1$ , we evaluate  $F(r)$  as

$$F(r) = (2t^3 - 3t^2 + 1)F(r_k) + (t^3 - 2t^2 + t)hr_k + (-2t^3 + 3t^2)F(r_{k+1}) + (t^3 - t^2)hm_{k+1},$$

where  $h = r_{k+1} - r_k$  and  $t = (r - r_k) / h$ .

## A.2 Generalized residuals

The CDF specification testing is based on the properties of the generalized residuals (also known as probability integral transform). First, for each out-of-sample period  $t = R + 1, \dots, T$ , we apply the CDF interpolation algorithm with input data  $(2r_{\min}, 0), (c_j, \widehat{\Pr}\{r_t \leq c_j | \mathcal{I}_{t-1}\}), (2r_{\max}, 1)$  for  $j = 1, \dots, p$ , where  $r_{\min}$  and  $r_{\max}$  are minimal and maximal sample values of returns within the estimation portion of the sample. That is, we approximate the conditional CDF values outside the interval  $[2r_{\min}, 2r_{\max}]$  by exact zero or exact one, which is reasonable, as the probability of such returns is negligible. The generalized residual  $\varepsilon_t$ ,  $t = R + 1, \dots, T$ , is computed simply as an interpolated conditional CDF evaluated at  $r_t$ .

## A.3 CDF Testing

The generalized residuals  $\varepsilon_t$  have a familiar property that  $\varepsilon_t \sim i.i.d. U[0, 1]$ . The univariate version of the generalized autocontours test of González-Rivera and Sun (2015) verifies whether the collection of  $k$  out-of-sample generalized residuals together with their lags are scattered uniformly inside the  $[0, 1]^{2k}$  hypercube.

The testing procedure contains the following steps. Let the vector  $\alpha$  contain  $p_\alpha$  ‘sides’  $\alpha_i \in (0, 1]$ , and consider pairs  $(\varepsilon_t, \varepsilon_{t-\ell})$  of out-of-sample generalized residuals and their  $\ell^{th}$  lags,  $\ell = 1, 2, \dots, L$ ,  $t = R + 1, \dots, T$ . Under the null hypothesis of correct specification

when  $\varepsilon_t \sim i.i.d. U[0, 1]$ , each side  $\alpha_i$  is consistently estimated by the sample proportion of pairs  $(\varepsilon_t, \varepsilon_{t-\ell})$  falling into the corresponding generalized autocontour – the subhypercube  $G\text{-ACR}_{\alpha, \ell} = \times_{i=1}^{p_\alpha} [0, \sqrt{\alpha_i}]^2$ :

$$\hat{\alpha}_{\alpha, \ell} = \frac{1}{T - R - \ell} \sum_{t=R+1+\ell}^T \mathbb{I}_{\{(\varepsilon_t, \varepsilon_{t-\ell}) \in G\text{-ACR}_{\alpha, \ell}\}}.$$

The González-Rivera–Sun test exists in two chi-squared variations: contour-aggregated and lag-aggregated. The contour-aggregated statistic gathers information from estimated generalized autocontours for a collection of different sides, keeping the lag, say  $\bar{\ell}$ , fixed. Let  $\hat{\alpha}_{\bar{\ell}} = (\hat{\alpha}_{1, \bar{\ell}}, \dots, \hat{\alpha}_{p, \bar{\ell}})'$ . The lag-aggregated statistic gathers information from estimated generalized autocontours for a collection of different lags, keeping the side, say  $\bar{\alpha}$ , fixed. Let  $\hat{\alpha}_{\bar{\alpha}} = (\hat{\alpha}_{\bar{\alpha}, 1}, \dots, \hat{\alpha}_{\bar{\alpha}, L})'$ .

Then, under the null of correct distributional specification,

$$GRS_{\alpha, \bar{\ell}} = (\hat{\alpha}_{\bar{\ell}} - \alpha)' A_{\alpha, \bar{\ell}}^{-1} (\hat{\alpha}_{\bar{\ell}} - \alpha) \rightarrow^d \chi_{p_\alpha}^2$$

and

$$GRS_{\bar{\alpha}, L} = (\hat{\alpha}_{\bar{\alpha}} - \bar{\alpha}\iota_L)' A_{\bar{\alpha}, L}^{-1} (\hat{\alpha}_{\bar{\alpha}} - \bar{\alpha}\iota_L) \rightarrow^d \chi_L^2,$$

where the matrices  $A_{\alpha, \bar{\ell}}$  and  $A_{\bar{\alpha}, L}$  contain asymptotic variances and covariances of elements of the estimated generalized autocontours, which are functions of elements of the vector  $\alpha$  only and need not be estimated (see González-Rivera and Sun (2015) for more details), and  $\iota_L$  is a column vector of ones of length  $L$ .

The rejection by the González-Rivera–Sun tests means that the generalized residuals are not likely to be uniform on  $[0, 1]$  and/or fail to be serially independent.

## A.4 Scoring rules

Gneiting and Raftery (2007) list several scoring rules that can be used to compare probabilistic forecasts of different models of (conditional) distributions. The Brier score for the forecast at  $t$  made at  $t - 1$  is

$$B_t = - \sum_{j=1}^{p+1} \left( \mathbb{I}_{\{c_{j-1} < r_t \leq c_j\}} - \widehat{\Pr}\{c_{j-1} < r_t \leq c_j\} \right)^2,$$

and is a quadratic criterion of deviations of binary realizations from probability forecasts. The CRPS for the forecast at  $t$  made at  $t - 1$  is

$$CRPS_t = - \int_{-\infty}^{\infty} \left( \widehat{\Pr}\{r_t \leq r | \mathcal{I}_{t-1}\} - \mathbb{I}_{\{r_t \leq r\}} \right)^2 dr,$$

where the conditional CDF  $\widehat{\Pr}\{r_t \leq r | \mathcal{I}_{t-1}\}$  is obtained by CDF interpolation (see Appendix A.1), while the integral is computed numerically using the Gauss-Chebyshev quadrature formulas (Judd (1998), section 7.2) with 300 Chebyshev quadrature nodes on  $[2r_{\min}, 2r_{\max}]$ .

The average Brier score and average CRPS are computed by averaging  $B_t$  and  $CRPS_t$  over out-of-sample periods  $t = R + 1, \dots, T$ .

## A.5 Trading strategy

To build a trading strategy, we use a simple rule exploring the difference between the predicted conditional probability and unconditional probability  $\Pr\{r_t \leq c_j\} = \alpha_j$ . We sum the differences over the interval of empirical quantiles  $[a, b]$  as

$$S_t = \sum_{c_j=a}^b \left( \widehat{\Pr}\{r_t \leq c_j | \mathcal{I}_{t-1}\} - \alpha_j \right).$$

If we sum all  $p$  available quantiles, we are using the information from the entire distribution. If we want to use the information only about positive returns, we sum only half of the available empirical quantiles corresponding to the cutoffs at positive returns. For example, in case the positive returns are predicted with a higher probability than the negative returns for all corresponding empirical quantiles, the sum  $S_t$  will be positive. Further, it may be useful to compare  $S_t$  computed for the empirical quantiles corresponding to both negative and positive returns. After some experimentation, we obtain threshold values for each stock depending on the shape of the conditional distributions, generating consistent profits. Hence we build the trading strategy on  $S_t$  exceeding these thresholds, but we note that this could further be optimized for maximum profits. In our setup, we use all quantiles, hence  $a = c_1$ , and  $b = c_p$ , while the threshold is set to zero.

Starting with a one dollar investment at the beginning of the sample, our investor decides to hold the stock depending on whether the predicted probability is favorable or not. We compare the cumulative returns from this simple market-timing strategy using predictions from the ordered logit, unordered logit, GARCH, FHS, and the buy and hold strategy for all 29 stocks separately.

## B Tables

Table 1. Estimates of intercepts  $\delta_{0,j}$  in the ordered logit specification  $\theta_{t,j} = \delta_{0,j} + x'_{t-1,j}\delta(\alpha_j)$  for the three illustrative stocks. The standard errors are below the point estimates.

$\alpha_j$	INTC	QCOM	XOM	$\alpha_j$	INTC	QCOM	XOM
5%	-3.046 (0.115)	-2.775 (0.116)	-2.871 (0.118)	95%	2.898 (0.155)	2.689 (0.113)	3.042 (0.122)
7.5%	-2.504 (0.096)	-2.422 (0.098)	-2.488 (0.102)	92.5%	2.704 (0.156)	2.581 (0.108)	2.579 (0.098)
10%	-2.195 (0.087)	-2.148 (0.088)	-2.235 (0.092)	90%	2.026 (0.095)	2.401 (0.100)	2.474 (0.098)
12.5%	-1.953 (0.081)	-1.999 (0.081)	-2.032 (0.088)	87.5%	1.702 (0.082)	2.137 (0.091)	2.105 (0.089)
15%	-1.800 (0.077)	-1.810 (0.075)	-1.812 (0.083)	85%	1.666 (0.079)	2.069 (0.089)	1.962 (0.087)
17.5%	-1.672 (0.074)	-1.684 (0.071)	-1.679 (0.082)	82.5%	1.455 (0.075)	1.957 (0.085)	1.751 (0.083)
20%	-1.489 (0.072)	-1.546 (0.068)	-1.575 (0.082)	80%	1.338 (0.072)	1.719 (0.080)	1.645 (0.084)
22.5%	-1.325 (0.070)	-1.397 (0.065)	-1.454 (0.082)	77.5%	1.214 (0.070)	1.528 (0.076)	1.492 (0.083)
25%	-1.151 (0.068)	-1.252 (0.064)	-1.301 (0.081)	75%	1.049 (0.066)	1.249 (0.068)	1.336 (0.081)
27.5%	-1.020 (0.067)	-1.074 (0.062)	-1.174 (0.081)	72.5%	0.899 (0.064)	1.100 (0.065)	1.171 (0.081)
30%	-0.859 (0.065)	-0.928 (0.061)	-1.033 (0.080)	70%	0.828 (0.063)	0.953 (0.062)	1.035 (0.080)
32.5%	-0.711 (0.064)	-0.781 (0.060)	-0.919 (0.080)	67.5%	0.764 (0.063)	0.831 (0.061)	0.873 (0.080)
35%	-0.631 (0.063)	-0.633 (0.059)	-0.813 (0.080)	65%	0.650 (0.062)	0.707 (0.060)	0.678 (0.079)
37.5%	-0.477 (0.062)	-0.542 (0.059)	-0.690 (0.080)	62.5%	0.539 (0.061)	0.622 (0.060)	0.531 (0.078)
40%	-0.386 (0.062)	-0.465 (0.059)	-0.595 (0.080)	60%	0.425 (0.061)	0.501 (0.059)	0.393 (0.079)
42.5%	-0.286 (0.062)	-0.344 (0.059)	-0.499 (0.080)	57.5%	0.334 (0.061)	0.354 (0.059)	0.227 (0.079)
45%	-0.223 (0.061)	-0.280 (0.059)	-0.345 (0.080)	55%	0.226 (0.061)	0.253 (0.059)	0.109 (0.079)
47.5%	-0.098 (0.062)	-0.134 (0.059)	-0.205 (0.079)	52.5%	0.096 (0.061)	0.144 (0.059)	0.023 (0.079)
0.50%	-0.004 (0.062)	-0.014 (0.059)	-0.117 (0.079)				

Table 2. Estimates of slope coefficients  $\kappa_{i,\ell}$  in the ordered logit specification  $\delta_\ell(\alpha_j) = \kappa_{0,\ell} + \sum_{i=1}^{q_\ell} 2^i(\alpha_j - 0.5)^i \cdot \kappa_{i,\ell}$  for the three illustrative stocks. The standard errors are below the point estimates.

Coefficient	INTC	QCOM	XOM	Coefficient	INTC	QCOM	XOM
$\kappa_{0,1}$	-0.003 (0.018)	-0.169 (0.015)	-0.053 (0.011)	$\kappa_{0,2}$	-3.33 (3.93)	-0.285 (4.001)	0.109 (8.104)
$\kappa_{1,1}$	-0.035 (0.044)	-0.143 (0.035)	-0.117 (0.043)	$\kappa_{1,2}$	4.92 (5.79)	-15.06 (3.08)	-17.52 (5.71)
$\kappa_{2,1}$	0.085 (0.084)	0.542 (0.092)	0.052 (0.076)	$\kappa_{2,2}$	12.13 (4.72)	-7.16 (5.61)	-16.85 (7.83)
				$\kappa_{3,2}$	-3.86 (7.32)	25.34 (7.524)	25.98 (9.23)

Table 3. Mean and median return-volatility characteristics from five trading strategies for all 29 stocks.

Method	mean			median		
	return	volatility	Sharpe	return	volatility	Sharpe
Ordered Logit	1.296	0.159	0.381	0.789	0.161	0.322
Separate Logits	1.088	0.159	0.289	0.663	0.155	0.289
Market	0.806	0.244	0.102	0.447	0.226	0.219
GARCH	0.768	0.202	0.169	0.414	0.182	0.277
FHS	0.791	0.193	0.162	0.428	0.184	0.225

## C Figures: Parameter Estimates

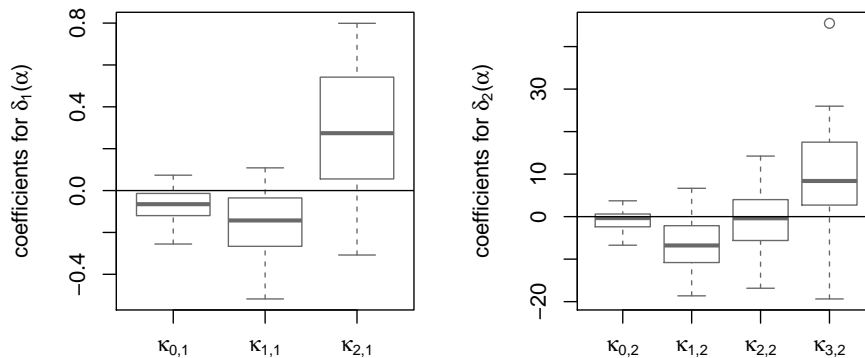


Figure 1: Parameter estimates: ordered logit parameters estimated for all 29 stocks shown by box-and-whisker plots.

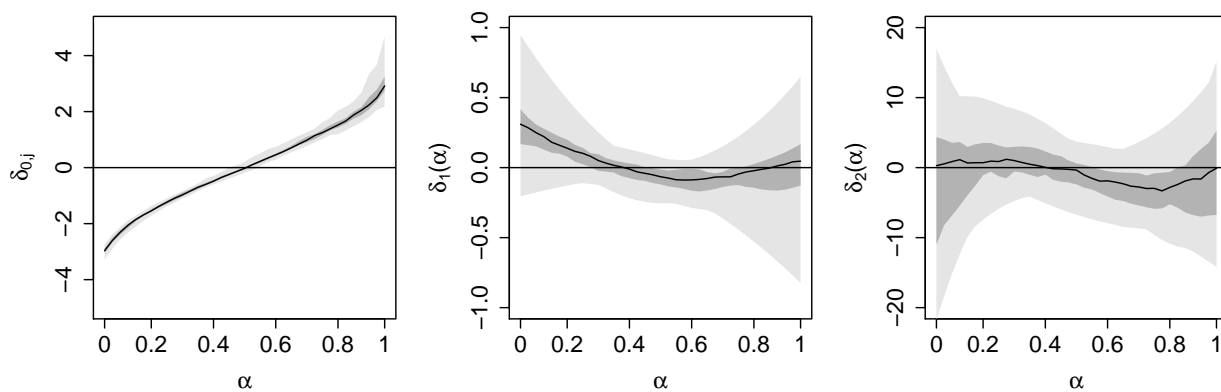


Figure 2: Parameter estimates: coefficient functions implied by parameters estimated for all 29 stocks. Minimum and maximum values are shown as a light grey area, 50% of the distribution as a grey area, and the median as a black line.

## D Figures: Conditional CDF

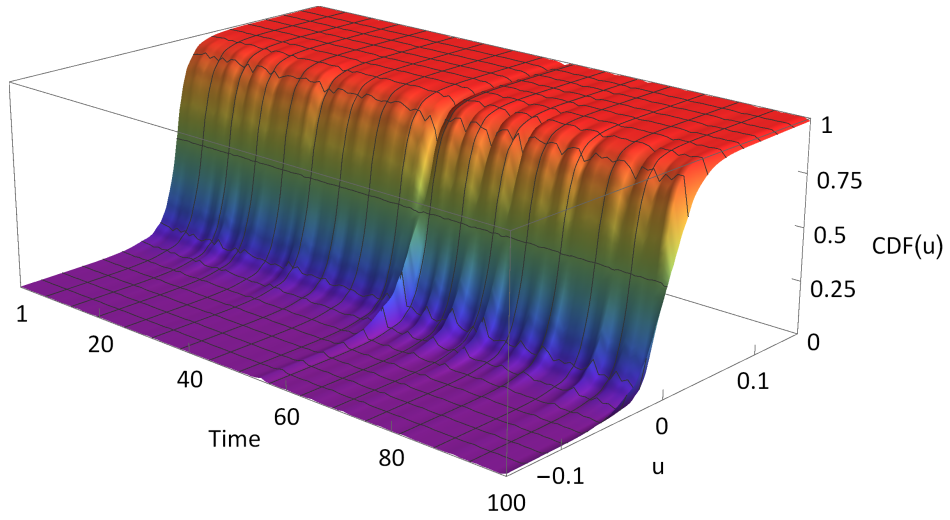


Figure 3: Fragment of interpolated estimated conditional CDF of returns for one of stocks.



## E Figures: Statistical Evaluation

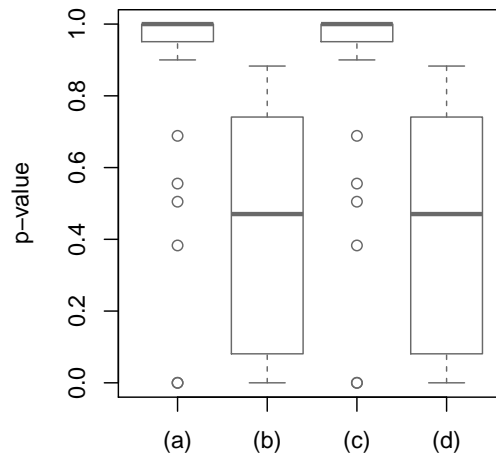


Figure 4: p-values from Gonzalez-Rivera and Sun (2015) tests for the ordered logit model shown by box-and-whisker plots for all 29 stocks. The four test specifications are shown with  $\alpha = (0.25, 0.5, 0.75)'$  (a) contour-aggregated, (b) lag-aggregated, with  $\alpha = (0.05, 0.1, \dots, 0.9, 0.95)'$  (c) contour-aggregated, (d) lag-aggregated.

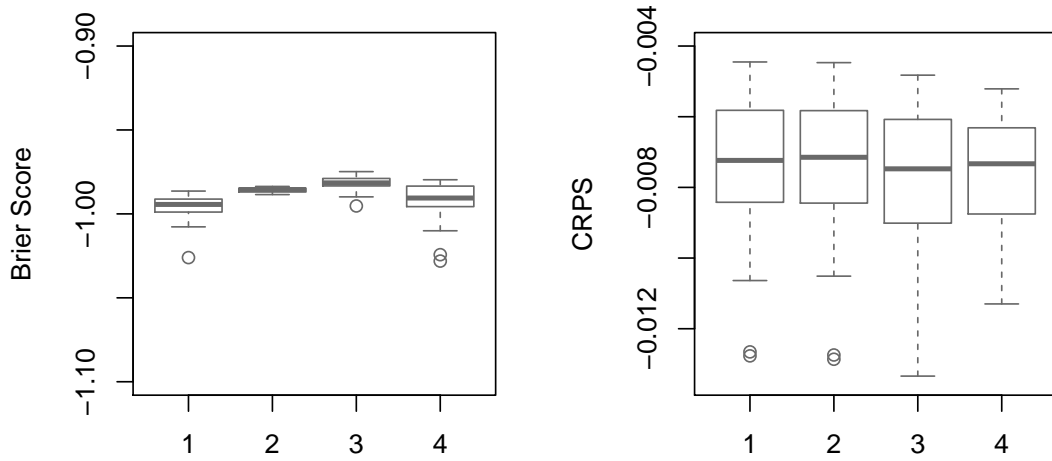


Figure 5: Brier scores and CRPS for ordered logit (1), separate logits (2), GARCH (3), and FHS (4), shown by box-and-whisker plots for all 29 stocks. Larger values of the scores are preferred to their smaller values.

## F Figures: Economic Evaluation

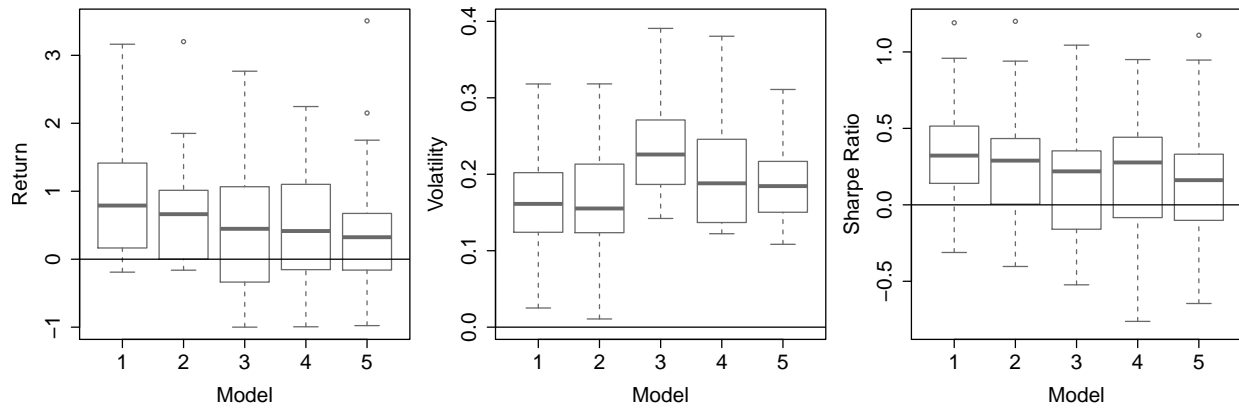


Figure 6: Performance of ordered logit (1), separate logits (2), market benchmark (3), GARCH (4), and FHS (5) models. Return, volatility and Sharpe ratio for all 29 stocks are shown by box-and-whisker plots.

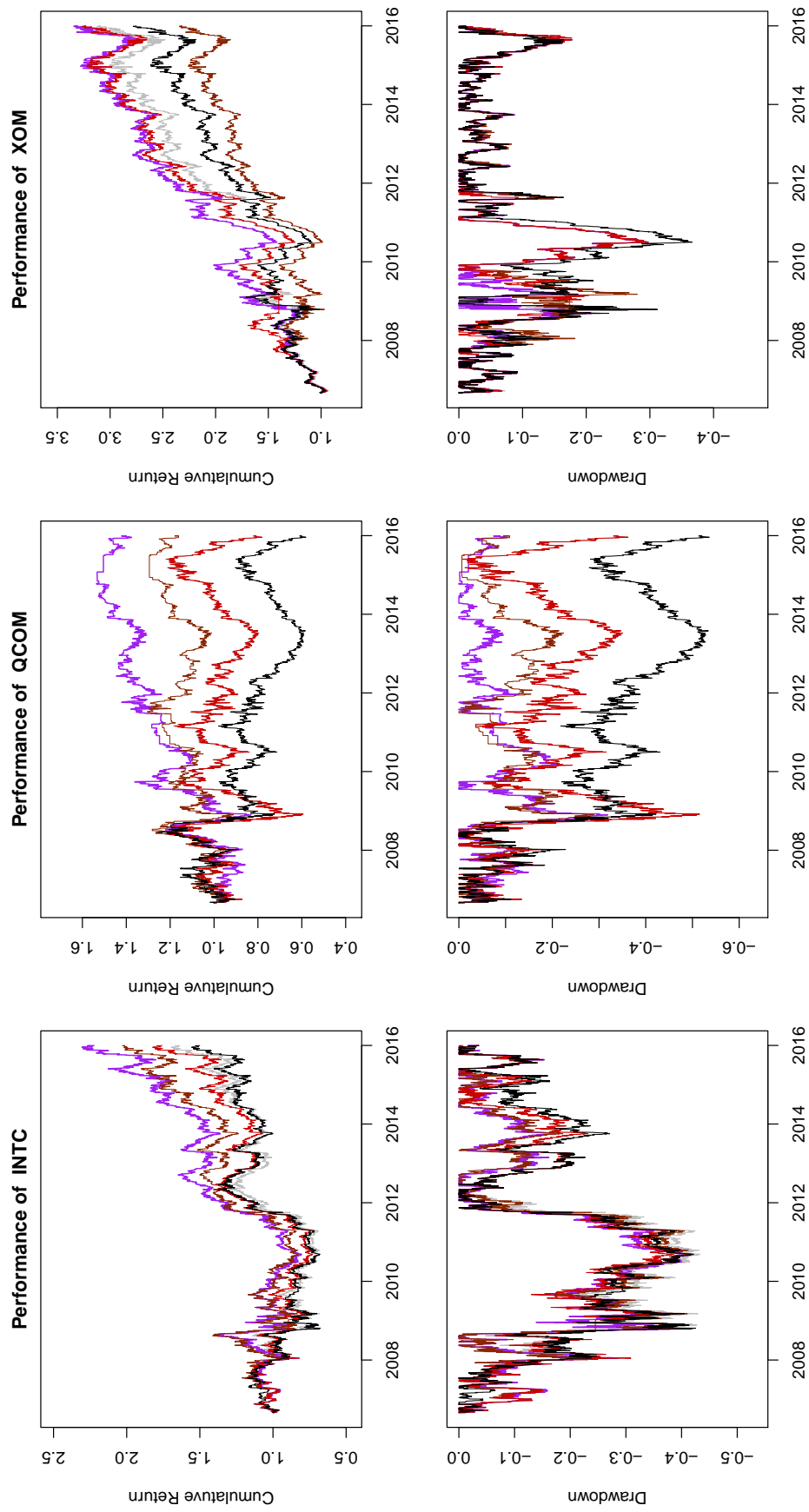


Figure 7: Performance: cumulative return and drawdown for three typical stocks. Trading strategy based on probability predictions from ordered logit model in purple (—), separate logits in red (—), GARCH in black (—), FHS in black (—), and benchmark buy and hold in grey (—).

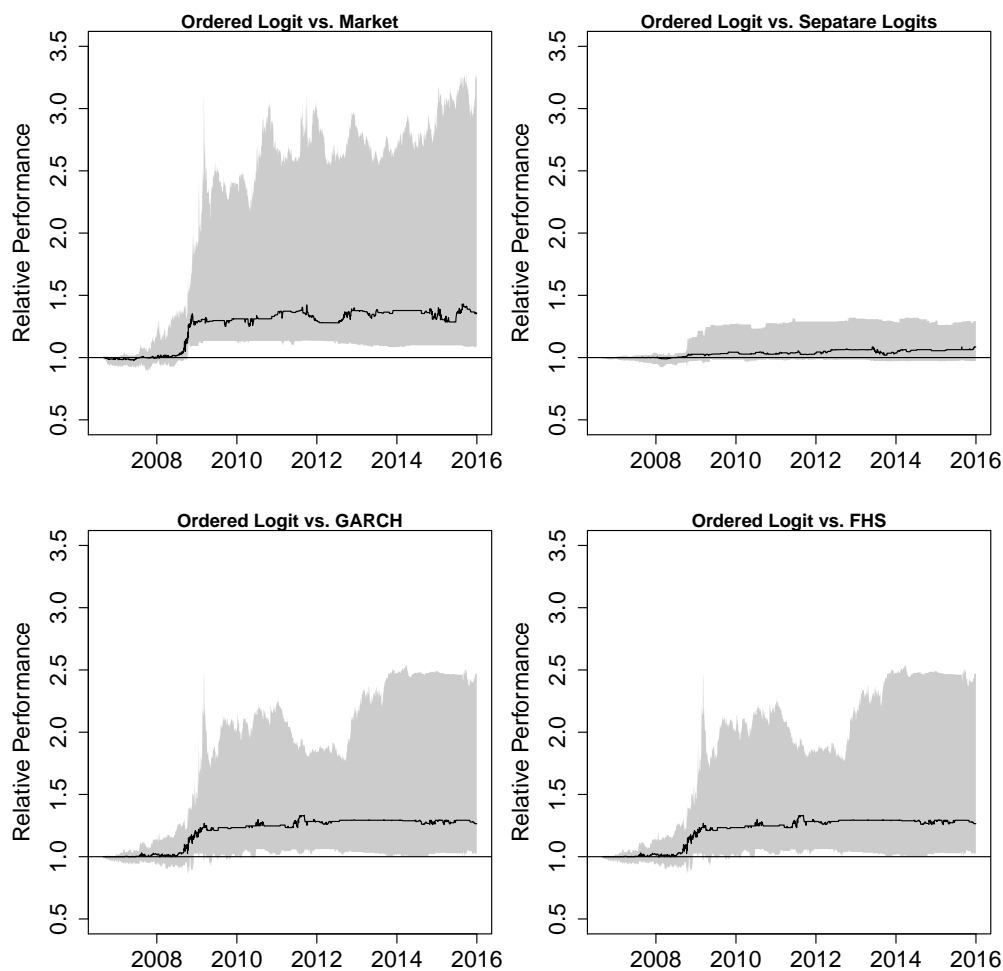


Figure 8: Relative performance: trading strategy based on probability predictions from the ordered logit model relative to separate logits (top right) as well as benchmark market (top left), GARCH (bottom left), and FHS (bottom right). The median value from all 29 stocks is black line surrounded by 90% of the distribution in grey area. Note that the value of one shows equal performance for both of the compared strategies.

## G Figures: Sensitivity to Polynomial Orders

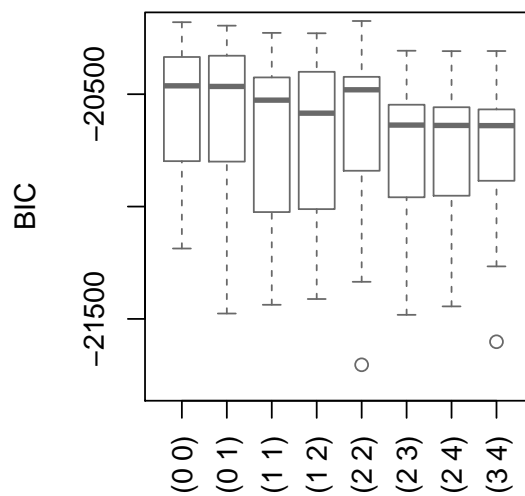


Figure 9: Bayesian information criterion for the ordered logit model with different polynomial orders  $(q_1, q_2)$  on the  $x$ -axis shown by box-and-whisker plots for all 29 stocks.