(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

International Journal of Forecasting 35 (2019) 823–835

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Forecasting dynamic return distributions based on ordered binary choice

Stanislav Anatolyev[a,b], Jozef Baruník[c,d,*]

[a] CERGE-EI, Politických Vězňů 7, 11121, Prague, Czech Republic
[b] New Economic School, 45 Skolkovskoe Shosse, Moscow, 121353, Russia
[c] Institute of Economic Studies, Charles University, Opletalova 26, 11000, Prague, Czech Republic
[d] The Czech Academy of Sciences, Institute of Information Theory and Automation, Pod Vodárenskou Věží 4, 182 08, Prague, Czech Republic

ARTICLE INFO

Keywords:
Asset returns
Predictive distribution
Conditional probability
Probability forecasting
Ordered binary choice

ABSTRACT

We present a simple approach to the forecasting of conditional probability distributions of asset returns. We work with a parsimonious specification of ordered binary choice regressions 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 a past volatility proxy as predictors. The 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 in which conditional distribution forecasts are desired.

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 have provided overwhelming evidence regarding the predictability of the first two moments of stock return distributions. The expected values of stock returns can be predicted to some extent using economic variables (Ang & Bekaert, 2006; Fama & French, 1989; Keim & Stambaugh, 1986; Viceira, 2012), while the conditional second moment can be characterized well by simple volatility models, or even measured from high-frequency data (Andersen, Bollerslev, Diebold, & Labys, 2003; Bollerslev, 1986). 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 only a small fraction of the literature.[1]

One of the main reasons why researchers do not tend to focus on characterizing the entire return distribution may be the 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 the von Neumann-Morgenstern expected utility. In contrast to this, Rostek (2010) recently developed a notion of quantile maximization and quantile utility

* Corresponding author at: Institute of Economic Studies, Charles University, Opletalova 26, 11000, Prague, Czech Republic.
E-mail address: barunik@fsv.cuni.cz (J. Baruník).

1 Few studies have focused on directional forecasts or threshold exceedances (Christoffersen & Diebold, 2006; Chung & Hong, 2007; Nyberg, 2011).

https://doi.org/10.1016/j.ijforecast.2019.01.005

preferences. This important shift in decision-theoretic foundations provokes us to depart from the limited mean–variance thinking and work with entire distributions.

The 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, and market timing strategies with precise entries and exits that reflect the information in the tails. Despite its importance, forecasting the conditional distribution of future returns has attracted little attention so far, in contrast to point forecasts and their uncertainty. This article presents a simple approach to forecasting a conditional distribution of stock returns using a parameterized ordered binary choice regression. While we focus here on financial returns, we note that our approach may be useful to many other applications in which the conditional distribution forecasts are of interest.

The majority of studies that focus on the prediction of conditional return distributions characterize the cumulative conditional distribution by a collection of conditional quantiles (Cenesizoglu & Timmermann, 2008; Engle & Manganelli, 2004; Pedersen, 2015; Žikeš & Baruník, 2016). 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. Foresi and Peracchi (1995) enable the approximation of the distribution function by estimating a sequence of conditional binary choice models over a grid of values that correspond to different points in 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 Chernozhukov, Fernández-Val, and Melly (2013), Fortin, Lemieux, and Firpo (2011), Hothorn, Kneib, and Bühlmann (2014), Rothe (2012), and Taylor and Yu (2016).

This article further develops the ideas set forth by Foresi and Peracchi (1995) and presents a simple related 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. We achieve this substantial reduction in the degree of parameterization by tying the coefficients of the predictors via a smooth dependence on corresponding probability levels. Our specification can be motivated in a semiparametric way, as we approximate smooth probability functions using low-order polynomials.

The probability forecasts are conditional on the past information contained in returns, as well as on their volatility proxy. The main reason for choosing the volatility as one of the explanatory variables is that the cross-sectional relationship between risk and expected returns, generally measuring a stock's risk as the covariance between its return and some factor, is documented well 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 regarding expected returns are essential for understanding classical asset pricing, little is known about the potential of these factors to identify extreme tail events of the return distribution precisely.

Our illustrative empirical analysis estimates conditional distributions of the 29 most liquid U.S. stocks and compares their generated forecasts with those from the buy-and-hold strategy and 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 from the collection of separate binary choice models. Section 3 contains information about the data we use and lays out the details of particular specifications. Section 4 presents empirical results, and Section 5 concludes. The appendix contains more technical material and details of some of the 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 the conditional cumulative return distribution $F(r_t \mid \mathcal{I}_{t-1})$ as precisely as possible, 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

$$c_1 < c_2 < \cdots < c_p,$$

and define $c_0 = -\infty$ and $c_{p+1} = +\infty$ for convenience. The higher p is, the more precise the description of the conditional distribution will be (a full discussion is provided later in this 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 α_j-quantile of returns, $j = 1, \dots, p$, where $0 < \alpha_1 < \alpha_2 < \cdots < \alpha_p < 1$ 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

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

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

Let $x_{t-1,j}, j = 1, \ldots, p$ be a vector of predictors for $\mathbb{I}_{\{r_t \le c_j\}}$ that may depend on j via the dependence of some of them on c_j. For instance, one of the predictors may be $\mathbb{I}_{\{r_{t-1} \le c_j\}}$, the past indicator (dependent on j), while another may be r_{t-1}, the past return (independent of j), and 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 is equal to k.

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

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

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, which leads to a substantial decrease in the degree of parameterization.

In the language of Foresi and Peracchi (1995), no monotonicity holds in general in the unordered model. That is, $\Pr\{r_t \le c_{j-1}|\mathcal{I}_{t-1}\}$ may exceed $\Pr\{r_t \le c_j|\mathcal{I}_{t-1}\}$ 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 not guaranteed to hold, but similar artificial measures can be applied. One simple way is to shift the value of a conditional distribution that violates monotonicity at a particular threshold to its value at the previous threshold plus an additional small amount. An alternative way to ensure both in-sample and out-of-sample predictability is via rearrangement (Chernozhukov, Fernández-Val, & Galichon, 2009). Given the generally low predictability of conditional probabilities for returns (and hence, their low variability compared to their mean), the share of observations that need such adjustments is expected to be low (see below for empirical evidence); thus, we give preference to the former, simpler method.

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

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

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

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

and $q_\ell \le p - 1$. Note that each intercept $\delta_{0,j}$ is j -specific and represents an 'individual effect' for a particular probability level, while the slopes' δs do not have index j; i.e., they depend on j only via dependence on the α_js. The motivation behind such a specification is semiparametric: any smooth function on [0, 1] can be approximated to a desired degree of precision by the system of basis polynomials $\{\alpha_j - 0.5, (\alpha_j - 0.5)^2, \ldots, (\alpha_j - 0.5)^q\}$ 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 in order to line up the coefficients κ_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 δ_j in each of p equations for θ_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 $\delta\left(\alpha_j\right)$, each parameterized via $1 + q_\ell$ parameters). The difference

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

grows with p, the fineness of the partition by thresholds. The resulting difference is also related positively to the number of predictors used.[2]

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

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

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

$$\ell_t^{(j)} = \mathbb{I}_{\{r_t \le c_j\}} \ln\left(\Lambda\left(\theta_{t,j}\right)\right), \tag{6}$$

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

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

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

2 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 $K_{UO} - K_O = 67$.

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, although not guaranteed, is easier to enforce in the 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 adjust automatically to these monotonicity restrictions. However, if the degrees of freedom are not sufficient to ensure this for all predictor values in the sample we can prevent (rare) realizations of negative differences for particular values of t by enforcing monotonicity through the constraints $\Delta_j \Lambda_t \geq \varepsilon$ for all j and t, where ε is some small number.[3]

Computationally, it is convenient to maximize the total likelihood in a number of steps, as an arbitrary initial parameter vector is likely to result in an incomputable likelihood, due to numerous violations of the monotonicity property. The idea is to begin by determining approximate values of individual intercepts and slopes, subject to their monotonicity, then relax 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[4]:

Step 1. Run a series of separate binary choice models as per Eqs. (1) and (2) with the specification $\theta_{t,j} = \delta_{0,j} + x'_{t-1,j}\delta_j$ with $\delta_j = \left(\delta_{j,1}, \ldots, \delta_{j,k}\right)'$, $j = 1, \ldots, p$, by maximizing the individual likelihoods in Eq. (6), and call the estimates thus obtained $\bar{\delta}_{0,j}$ and $\bar{\delta}_j$.

Step 2. For each $\ell = 1, \ldots, k$, 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 $\delta_\ell = \left(\delta_{1,\ell}, \ldots, \delta_{p,\ell}\right)'$ and $\alpha = \left(\alpha_1, \ldots \alpha_p\right)'$, and call the estimates thus obtained $\bar{\kappa}_{i,\ell}$, $i = 0, 1, \ldots, q_\ell$.

Step 3. Run the ordered binary choice model in Eqs. (1) and (3) with the specification in Eq. (4) by maximizing the total likelihood in Eq. (7) using $\bar{\delta}_{0,j}$ as starting points for $\delta_{0,j}$, $j = 1, \ldots, p$, and $\bar{\kappa}_{i,\ell}$ as starting points for $\kappa_{i,\ell}$, $i = 0, 1, \ldots, q_\ell$, $\ell = 1, \ldots, k$.

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 the monotonicity of the outcome. To this end, we apply the Fritsch–Carlson monotonic cubic interpolation (Fritsch & Carlson, 1980, see also 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 the approximation of the conditional distribution constructed in the ordered model is determined by a number of factors, with 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 δ on the α_js. It seems reasonable to set the system of thresholds to be fine enough (as long as one does not come close to the computability limits) to describe the distribution with sufficient precision, but not so fine that a considerable number of observations fall between each pair of adjacent thresholds. One may also afford higher flexibility to the slope specification for larger sample sizes; in practice, though, low numbers are usually pretty adequate in semiparametric setups. One may also employ formal model selection criteria such as the Bayesian information criterion for choosing the optimal orders of polynomials in slope specifications.

3. Data and empirical specification

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

Next we provide 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[6] from $\alpha \in (5\%, 95\%)$, i.e., a grid with step 2.5% resulting in a total of $p = 37$ quantiles.[7] We use a time-varying

---

3 In our empirical illustration, we forced each difference to be bounded below by $\varepsilon = 10^{-6}$. Such a strategy has resulted in under 1% of such interferences among all differences during in-sample estimation and under 2% during out-of-sample forecasting.

4 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.

5 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), and Exxon Mobil Corporation (XOM).

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

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

partition that changes with the rolling window. In each window, c_j is computed as $\gamma(\alpha_j)\sqrt{\sigma_t^2}$, where $\gamma(\alpha_j)$ is a quantile of the standard normal distribution, and σ_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 $j = 1, \ldots, p$. The first predictor is a lagged indicator that corresponds to the probability level α_j, which is supposed plausibly to have the highest predictive 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 other hand, all predictors may vary with the quantile. In the ordered model, after some experimentation with statistical significance of higher-order polynomials, we set $q_1 = 2$ and $q_2 = 3$.[8] That is, the polynomial is quadratic in the probability level α for the past indicator and cubic for the past volatility proxy.

The full specification of the model for $j = 1, \ldots, p$ 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

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

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

There are a total of $K_O = 44$ 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 sufficiently parsimonious and approximates the distribution quite well, and additional terms do not bring significant improvements. Hence, estimating seven parameters $\kappa_{i,\ell}$ in addition to the 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. As we consider forecasts for 29 stocks, we begin by presenting individual estimates of three illustrative stocks, namely Intel Corporation (INTC), QUALCOMM, Inc. (QCOM), and Exxon Mobil Corporation (XOM), then present the results for all 29 stocks in an aggregate form.

After presenting the parameter estimates, we evaluate the statistical and economic significance of the predicted distributions. Furthermore, we compare the performance of the ordered model with those of popular and challenging benchmarks used in the literature. Specifically, we include two candidate models that Kuester, Mittnik, and Paolella (2006) found to perform best: an asymmetric generalized autoregressive conditional heteroscedastic model with a skewed-t distribution (GARCH henceforth) and a GARCH-filtered historical simulation (FHS henceforth). Our implementation of these two alternative methods follows 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 the expected signs — negative for probability levels to the left of 50% and positive to the right of 50% — due to a monotonically increasing $\Lambda(\cdot)$ and a low predictability of the predictors, and exhibit the 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 a few cutoff points near the median. The intercept values are quite similar across the three stocks, though there is some variability.

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

The estimates of the seven coefficients $\kappa_{i,\ell}$ that drive 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 small in magnitude, and some are highly statistically significant. The left plot in Fig. 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, namely the past volatility proxy $\ln(1 + |r_{t-1}|)$, carries even more important information about future probabilities. The estimated coefficients $\kappa_{i,2}$ that correspond to the volatility predictor reveal that the cubic polynomial has many statistically significant coefficients (see Table 2). The right plot of Fig. 1 shows estimated coefficients for all 29 stocks as box-and-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 the return distributions.

[8] We also perform a model selection analysis with the Bayesian information criterion in the ordered model; see Section 4.4 for details.

Fig. 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 Fig. 2). Overall, we can see that the effect is small. The coefficient function $\delta_2(\alpha_j)$ implied by the past volatility proxy shows a similar impact.

Fig. 3 depicts an interpolated predictive conditional CDF of returns on the interval $[-1.5, 1.5]$ for an arbitrary stock[9] evaluated at arbitrary 100 out-of-sample periods. This allows one to observe how the cumulative distribution varies over time. There is a certain distribution clustering, i.e., the CDF possesses some persistence, while at other periods the CDF shape stands out from the cluster.

4.2. Statistical fit

We assess the adequacy of the predicted distribution by running the generalized autocontours specification test of González-Rivera and Sun (2015). This test verifies whether the collection of out-of-sample generalized residuals (also known as the probability integral transform), together with their lags, are scattered uniformly inside a hypercube of corresponding dimensions; 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 $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_p)'$. We use a side $\bar{\alpha} = 0.5$ in the lag-aggregated test with $L = 3$ and $L = 10$. Fig. 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.[10]

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

4.3. Economic significance

We investigate the economic usefulness of the proposed model by studying a simple profit rule for timing the market based on the model forecasts. The idea is to explore information from the entire distribution. We build a trading strategy by exploring the difference between the predicted conditional and unconditional probabilities, which indicates whether positive or negative returns are predicted with a higher probability. In the spirit of the previous literature (Anatolyev & Gospodinov, 2010; Breen, Glosten, & Jagannathan, 1989; Pesaran & Timmermann, 1995), 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 mean and median returns and the volatility, as well as the Sharpe ratio. While the benchmark market strategy earns a 0.806 return with a 0.244 volatility, GARCH and FHS generate similar returns with lower volatilities. When an investor uses predictions about the entire distribution from separate logit models, one obtains an improved mean return of 1.088 with a much lower volatility. Finally, our proposed ordered logit model generates a 1.296 mean return with a volatility that is similar to those of separate logits. The Sharpe ratio showing a risk-adjusted return, or an 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.

Fig. 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, with these returns being quite heterogeneous for all stocks considered. GARCH and FHS improve the volatility estimates, and hence yield similar returns with a lower risk. In terms of risk-adjusted Sharpe ratios, the separate logistic regressions yield similar results, while the ordered logit model shows a marked improvement. One can see positive Sharpe ratios for almost all stocks considered, pointing to the highest risk-adjusted returns offered by the ordered model.

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

Finally, Fig. 8 compares the relative performances of all five strategies. 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 Fig. 8 compare the ordered logit with the GARCH and FHS models. We document consistently better performances for the proposed ordered logit model than for either the unordered logit or benchmark strategies for all 29 stocks.

4.4. Sensitivity analysis

As has been noted, after some experimentation we have chosen a specification with $p = 37$ thresholds

[9] The minimal and maximal values of returns in the whole sample for this stock are approximately -0.20 and 0.20.

[10] We do not account for the estimation noise when constructing the test, as this generally tends to increase asymptotic variances, meaning that the p-values would be even higher if the estimation noise was accounted for.

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

and polynomial orders of $q_1 = 2$ and $q_2 = 3$. Here, we report on the results with different specifications. As a robustness check, we conduct 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 three 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 a value of p may have a serious impact on the model complexity: while the number of parameters for the basic partition $p = 37$ is equal to $K_O = 44$, that for the finer partition is equal to 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 $p = 73$,[12] some of the generalized autocontour tests (in particular, the three-side contour-aggregated test) exhibit the tendency to reject the constructed conditional distribution. However, the results of the economic evaluation are not sensitive to the choice of partition, and the dominance of the proposed method based on the ordered logit over all of the other methods still prevails.

On the one hand, the polynomial orders (q_1, q_2) do not have a drastic effect on the degree of parsimony; on the other hand, the dependence of predictability on the probability level is presumably not so sophisticated as to require higher-order powers. Hence, one would not expect a 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 the BIC is presented in Fig. 9. While the values of q_1 and q_2 such as zero and one 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 the 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 the conditional probability distribution of returns well. We subject the proposed model to a number of statistical tests for adequacy, and to comparisons with alternative methods. In order to see how useful the model is economically, we use distributional predictions in a simple market timing strategy. Using 29 liquid U.S. stocks, we find significant economic gains relative to the benchmarks.

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

[12] The detailed results are available from the authors upon request.

Acknowledgments

We are grateful to the editor Dick van Dijk and to two anonymous referees for useful comments and suggestions, which have improved the article greatly. Support from the Czech Science Foundation (GAČR) under the GA16-14179S project is gratefully acknowledged.

Appendix A

A.1. CDF interpolation

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

Suppose that 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 the construction of the estimated distribution. We start by computing the slopes of the secant lines as $\Delta_k = (F(r_{k+1}) - F(r_k))/(r_{k+1} - r_k)$ for $k = 1, \dots, K-1$, then compute 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 the probability integral transform). First, for each out-of-sample period $t = R+1, \dots, T$, we apply the CDF interpolation algorithm with input data $(2r_{\min}, 0), (c_j, \widehat{\Pr}\{r_t \le c_j | \mathcal{I}_{t-1}\}), (2r_{\max}, 1)$ for $j = 1, \dots, p$, where $r_{\min}$ and $r_{\max}$ are the 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 exactly zero or exactly one, which is reasonable, as the probability of such returns is negligible. The generalized residual ε_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 ε_t have the 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 and their lags are scattered uniformly inside the $[0, 1]^{2k}$ hypercube.

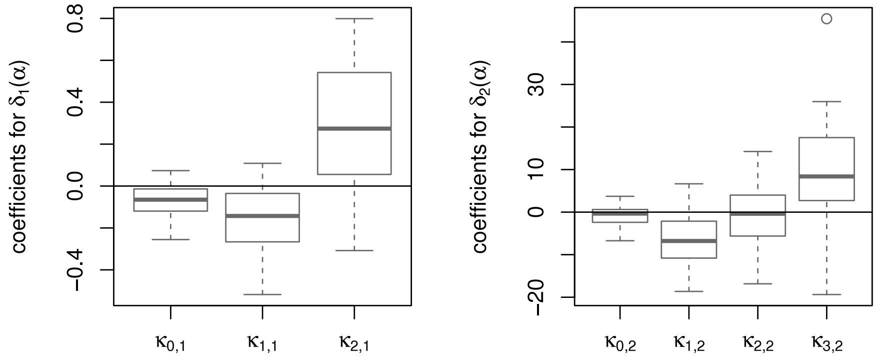

Fig. 1. Parameter estimates: ordered logit parameters estimated for all 29 stocks, shown as box-and-whisker plots.

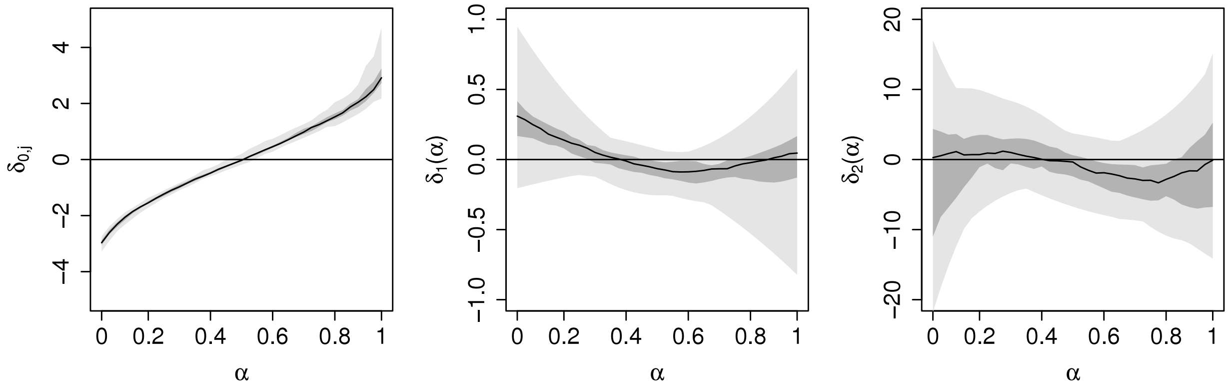

Fig. 2. Parameter estimates: coefficient functions implied by the parameters estimated for all 29 stocks. Minimum and maximum values are shown as a light grey area, 50% of the distribution is shown as a grey area, and the median is shown as a black line.

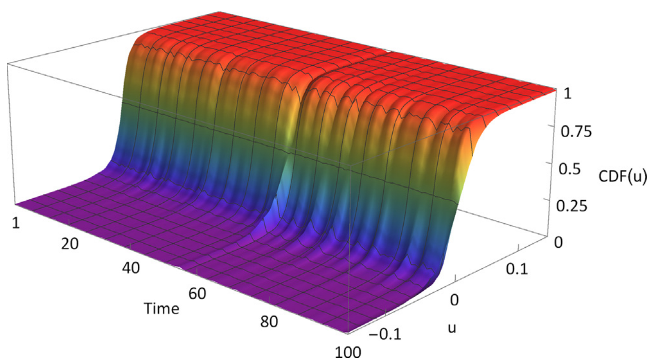

Fig. 3. Fragment of the interpolated estimated conditional CDF of returns for one of the stocks.

The testing procedure consists of the following steps. Let the vector α contain p_α 'sides' $\alpha_i \in (0, 1]$, and consider pairs $(\varepsilon_t, \varepsilon_{t-\ell})$ of out-of-sample generalized residuals and their ℓth lags, $\ell = 1, 2, \ldots, L$, $t = R+1, \ldots, T$. Under the null hypothesis of correct specification when $\varepsilon_t \sim i.i.d.\, U[0, 1]$, each side α_i is estimated consistently by the sample proportion of pairs $(\varepsilon_t, \varepsilon_{t-\ell})$ that fall into the corresponding generalized autocontour — the subhypercube G-ACR$_{\alpha,\ell} = \times_{i=1}^{p_\alpha}[0, \sqrt{\alpha_i}]^2$:

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

The González-Rivera–Sun test exists in two chi-squared variations: contour-aggregated and lag-aggregated. The

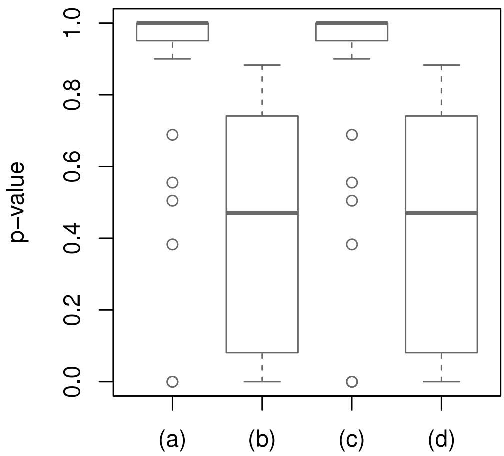

Fig. 4. p-values from González-Rivera and Sun (2015) tests for the ordered logit model for all 29 stocks, shown as box-and-whisker plots. The four test specifications are shown with $\alpha = (0.25, 0.5, 0.75)'$ (a) contour-aggregated and (b) lag-aggregated; and with $\alpha = (0.05, 0.1, \ldots, 0.9,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

contour-aggregated statistic gathers information from estimated generalized autocontours for a collection of different sides, keeping the lag, say $\bar{\ell}$, fixed. Let $\hat{\alpha}_{\bar{\ell}} = (\hat{\alpha}_{1,\bar{\ell}}, \ldots, \hat{\alpha}_{p,\bar{\ell}})'$. The lag-aggregated statistic gathers information from estimated generalized autocontours for a collection of different lags, keeping the side, say $\bar{\alpha}$, fixed. Let $\hat{\alpha}_{\bar{\alpha}} = (\hat{\alpha}_{\bar{\alpha},1}, \ldots, \hat{\alpha}_{\bar{\alpha},L})'$.

Then, under the null of correct distributional specification,

$$GRS_{\alpha,\bar{\ell}} = (\hat{\alpha}_{\bar{\ell}} - \alpha)' A^{-1}_{\alpha,\bar{\ell}} (\hat{\alpha}_{\bar{\ell}} - \alpha) \to^d \chi^2_{p_\alpha}$$

and

$$GRS_{\bar{\alpha},L} = (\hat{\alpha}_{\bar{\alpha}} - \bar{\alpha}\iota_L)' A^{-1}_{\bar{\alpha},L} (\hat{\alpha}_{\bar{\alpha}} - \bar{\alpha}\iota_L) \to^d \chi^2_L,$$

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

A 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 for t made at $t-1$ is

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

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

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

where the conditional CDF $\widehat{\Pr}\{r_t \le r|\mathcal{I}_{t-1}\}$ 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 Chebychev quadrature nodes on $[2r_{\min}, 2r_{\max}]$.

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

A.5. Trading strategy

We build a trading strategy by using a simple rule that explores the difference between the predicted conditional probability and the unconditional probability $\Pr\{r_t \le c_j\} = \alpha_j$. We sum the differences over the interval of empirical quantiles $[a, b]$ as

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

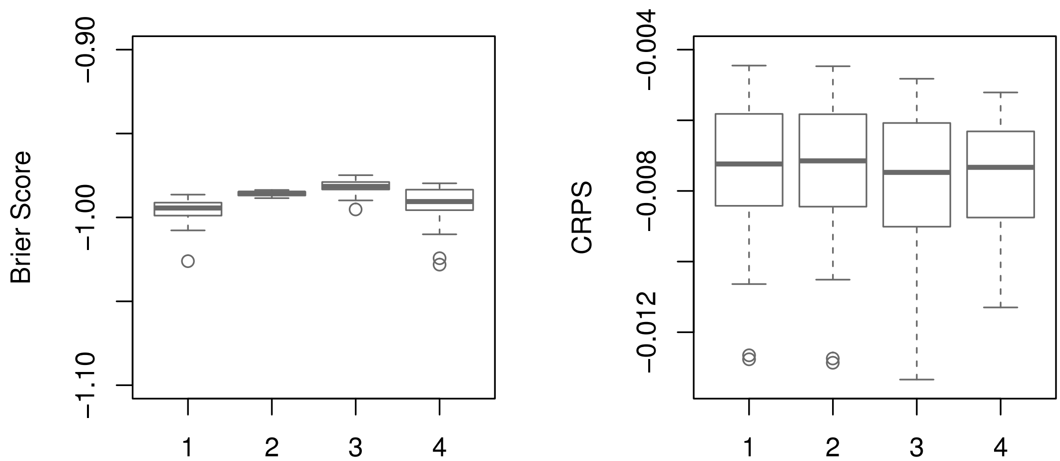

Fig. 5. Brier scores and CRPS for (1) ordered logit, (2) separate logits, (3) GARCH, and (4) FHS, shown as box-and-whisker plots for all 29 stocks. Larger values of the scores are preferred to smaller values.

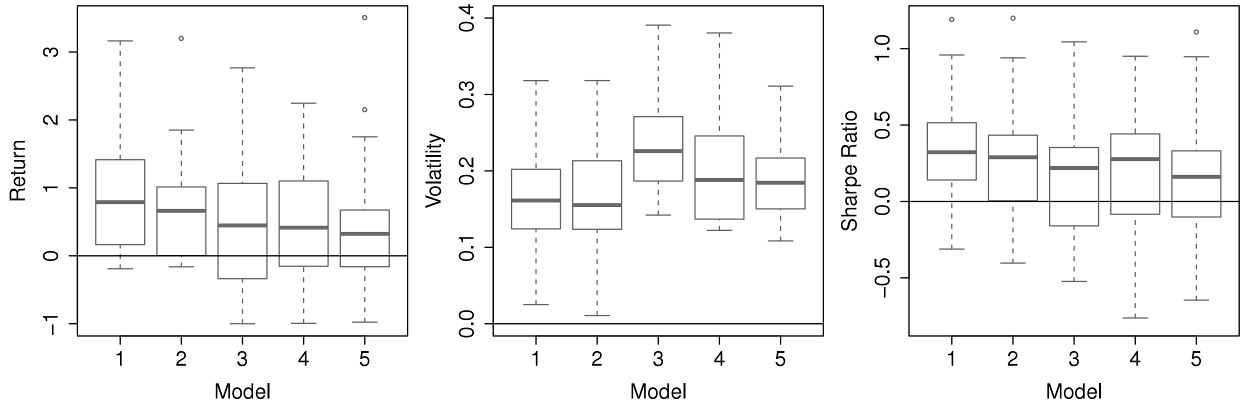

Fig. 6. Performances of (1) ordered logit, (2) separate logits, (3) market benchmark, (4) GARCH, and (5) FHS models. Returns, volatilities and Sharpe ratios for all 29 stocks are shown as box-and-whisker plots.

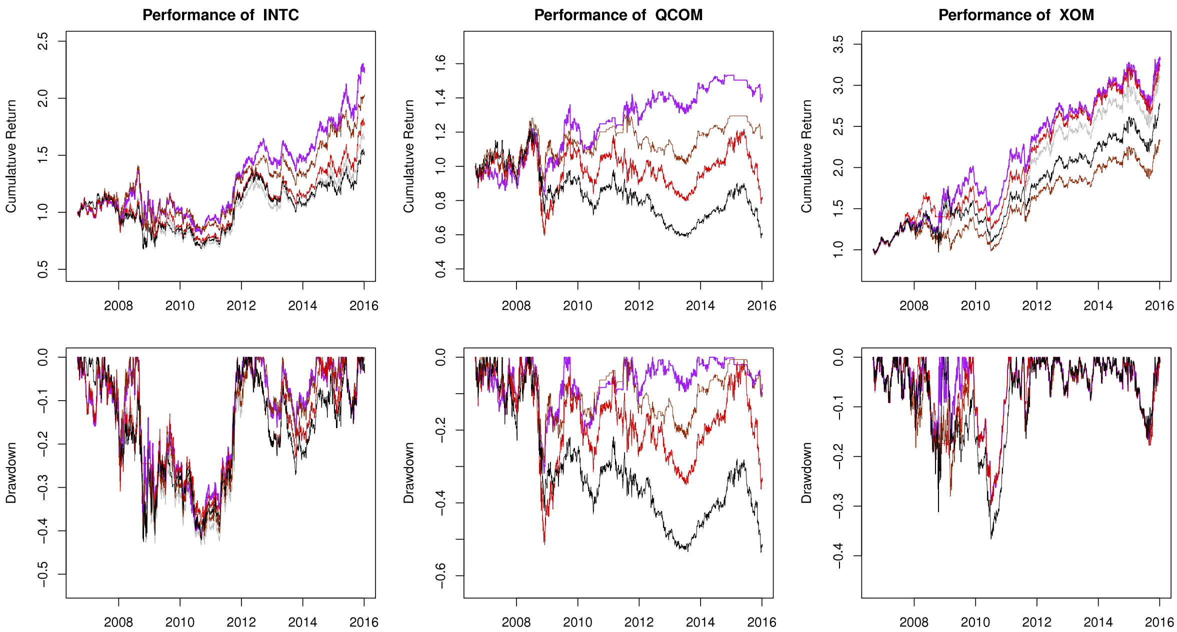

Fig. 7. Performance: cumulative returns and drawdowns for three typical stocks. A trading strategy based on probability predictions from the ordered logit model is shown in purple (–), separate logits are shown in bordeaux (–), GARCH is shown in red (–), FHS is shown in black (–), and the benchmark buy and hold is shown in grey (–). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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

Starting with a \$1 investment at the beginning of the sample, our investor decides to hold the stock based 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, the unordered logit, GARCH, FHS, and the buy and hold strategy for all 29 stocks separately.

Appendix B. Tables

See Tables 1–3.

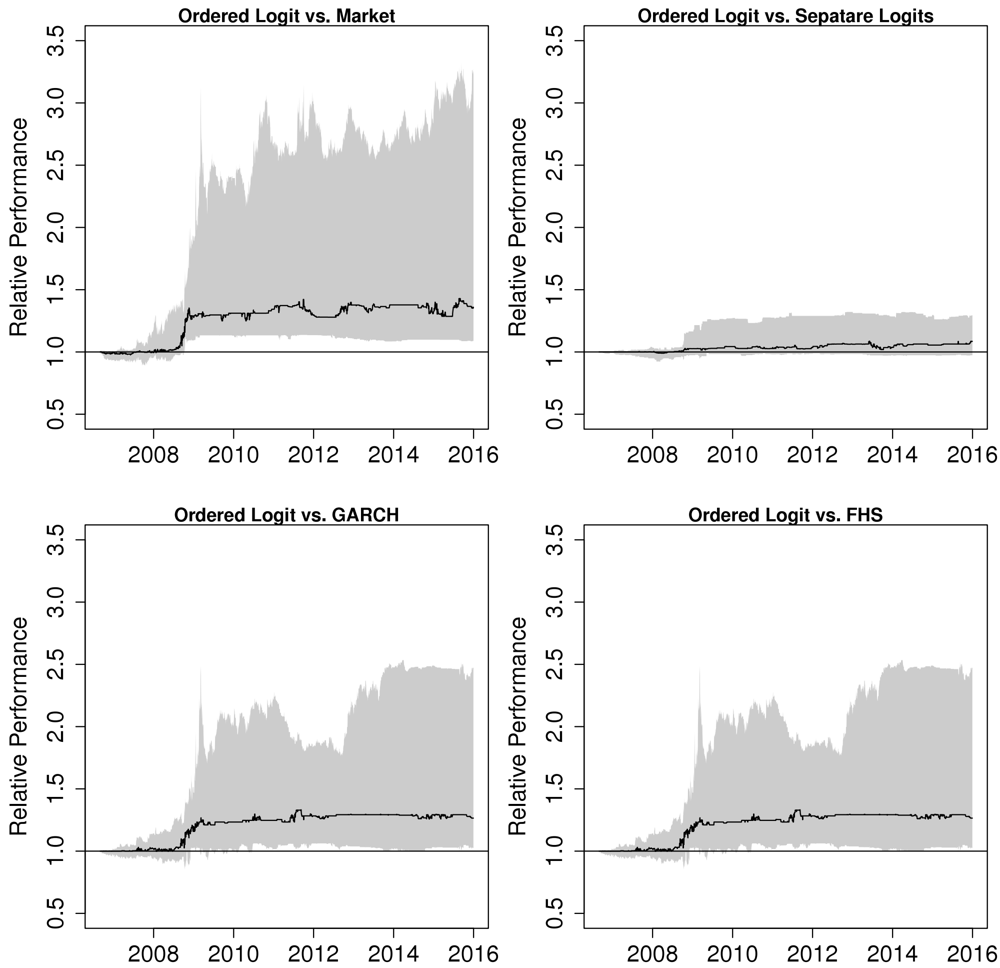

Fig. 8. Relative performances: a trading strategy based on probability predictions from the ordered logit model relative to separate logits (top right), as well as the benchmark market (top left), GARCH (bottom left), and FHS (bottom right). The median value from all 29 stocks is the black line, surrounded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

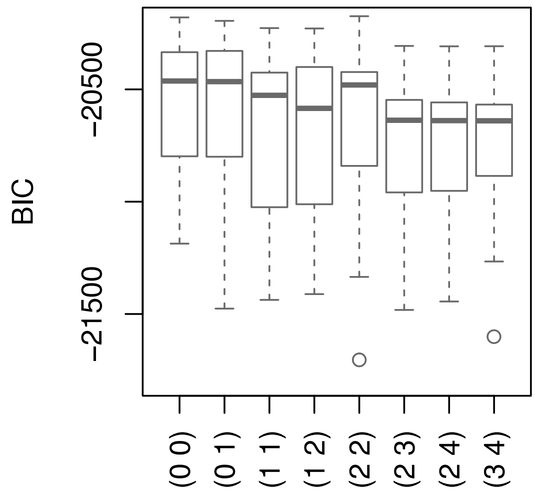

Fig. 9. Bayesian information criteria for the ordered logit model with different polynomial orders $(q_1 q_2)$ on the x-axis, shown as box-and-whisker plots for all 29 stocks.

Appendix C. Figures: Parameter estimates

See Figs. 1 and 2.

Appendix D. Figures: Conditional CDF

See Fig. 3.

Appendix E. Figures: Statistical evaluation

See Figs. 4 and 5.

Appendix F. Figures: Economic evaluation

See Figs. 6–8.

Appendix G. Figures: Sensitivity to polynomial orders

See Fig. 9.

Table 1
Estimates of the intercepts $\delta_{0,j}$ in the ordered logit specification $\theta_{t,j} = \delta_{0,j} + x'_{t-1,j}\delta\left(\alpha_j\right)$ for the three illustrative stocks.

α_j	INTC	QCOM	XOM	α_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)
50%	−0.004 (0.062)	−0.014 (0.059)	−0.117 (0.079)				

Note: Standard errors are given below the point estimates.

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

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)	−0.29 (4.00)	0.11 (8.10)
$\kappa_{1,1}$	−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 (7.52)	25.98 (9.23)

Note: Standard errors are given below the point estimates.

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

References

Anatolyev, S., & Gospodinov, N. (2010). Modeling financial return dynamics via decomposition. *Journal of Business & Economic Statistics*, *28*(2), 232–245.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, *71*(2), 579–625.

Ang, A., & Bekaert, G. (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., Glosten, L. R., & Jagannathan, R. (1989). Economic significance of predictable variations in stock index returns. *Journal of Finance*, *44*(5), 1177–1189.

Cenesizoglu, T., & Timmermann, A. G. (2008). Is the distribution of stock returns predictable? Manuscript, UCSD.

Chernozhukov, V., Fernández-Val, I., & Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, *96*(3), 559–575.

Chernozhukov, V., Fernández-Val, I., & Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, *81*(6), 2205–2268.

Christoffersen, P. F., & Diebold, F. X. (2006). Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science*, *52*(8), 1273–1287.

Chung, J., & Hong, Y. (2007). Model-free evaluation of directional predictability in foreign exchange markets. *Journal of Applied Econometrics*, *22*(5), 855–889.

Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, *22*(4), 367–381.

Fama, E. F., & French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economic*, *25*(1), 23–49.

Foresi, S., & Peracchi, F. (1995). The conditional distribution of excess returns: An empirical analysis. *Journal of the American Statistical Association*, *90*(430), 451–466.

Fortin, N., Lemieux, T., & Firpo, S. (2011). Decomposition methods in economics. In *Handbook of Labor Economics, vol. 4* (pp. 1–102). Elsevier.

Fritsch, F. N., & Carlson, R. E. (1980). Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, *17*(2).

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378.

González-Rivera, G., & Sun, Y. (2015). Generalized autocontours: Evaluation of multivariate density models. *International Journal of Forecasting*, *31*(3), 799–814.

Hothorn, T., Kneib, T., & Bühlmann, P. (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., & Stambaugh, R. F. (1986). Predicting returns in the stock and bond markets. *Journal of Financial Economic*, *17*(2), 357–390.

Kuester, K., Mittnik, S., & Paolella, M. S. (2006). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Economic*, *4*(1), 53–89.

Leorato, S., & Peracchi, F. (2015). *Comparing Distribution and Quantile Regression*: *Tech. Rep.*, 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., & Timmermann, A. (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., & Yu, K. (2016). Using autoregressive 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., & Baruník, J. (2016). Semi-parametric conditional quantile models for financial returns and realized volatility. *Journal of Financial Economic*, *14*(1), 185–226.