Centre for
Economic
and Financial
Research
at
New Economic
School

November 2009

Directional Prediction of Returns under Asymmetric Loss: Direct and Indirect Approaches

Stanislav Anatolyev
Natalia Kryzhanovskaya

Working Paper No 136

CEFIR / NES Working Paper series

Directional prediction of returns under asymmetric loss: direct and indirect approaches

Stanislav Anatolyev* New Economic School

Natalia Kryzhanovskaya New Economic School

November 2009

Abstract

To predict a return characteristic, one may construct models of different complexity describing the dynamics of different objects. The most complex object is the entire predictive density, while the least complex is the characteristic whose forecast is of interest. This paper investigates, using experiments with real data, the relation between the complexity of the modeled object and the predictive quality of the return characteristic of interest, in the case when this characteristic is a return sign, or, equivalently, the direction-of-change. Importantly, we carry out the comparisons assuming that the underlying loss function is asymmetric, which is more plausible than the quadratic loss still prevailing in the analysis of returns. Our experiments are performed with returns of various frequencies on a stock market index and exchange rate. By and large, modeling the dynamics of returns by autoregressive conditional quantiles tends to produce forecasts of higher quality than modeling the whole predictive density or modeling the return indicators themselves.

Key words: Directional prediction, sign prediction, model complexity, prediction quality, asymmetric loss, predictive density, conditional quantiles, binary autoregression.

*Corresponding author. Address: Stanislav Anatolyev, New Economic School, Nakhimovsky Pr., 47, office 1721(3), Moscow, 117418 Russia. E-mail: sanatoly@nes.ru.

1 Introduction

One of important choices in making model-based time-series predictions is the degree of complexity of the object whose dynamics will be modeled. On the one hand, one may model the whole one-period ahead conditional density, and read off the forecast for the characteristic of interest from the estimated conditional density. This is an *indirect approach* to forecasting. It has an advantage that it contains all the information about the dynamics of the variable of interest and hence may be used for a variety of purposes, while its obvious shortcoming is difficult implementation, in particular, of adequate specification and precise estimation. On the other hand, one may model the dynamics of the feature of interest in the first place, which gives rise to the *direct approach*. It is easier to implement and it contains minimally necessary information,[1] but the model may not be sufficiently flexible. Intermediate, *semi-direct approaches* are also possible. There, one describes the evolution of an object (usually more complex, but not necessarily) different from the characteristic of interest which is however simpler than the whole conditional density.

Here are three examples of direct vs. indirect (and possibly semi-direct) approaches. In the value-at-risk (VaR) literature, the indirect approach would be to analyze the conditional distribution (Kuester, Mittnik and Paolella, 2006), the semi-direct one would be to focus on its tails (e.g., McNeil and Frey, 2000), and the direct approach would be to parameterize the evolution of a conditional quantile (Engle and Manganelli, 2004). Kuester, Mittnik and Paolella (2006) recently showed that the quantile models tend to be inferior to fully parametric models in terms of forecasting ability. Another example is prediction under the linear exponential loss. While most of the literature tends to exploit the indirect approach (e.g., Patton and Timmermann, 2007) or use approximations (e.g., Christoffersen and Diebold, 1997), Anatolyev (2009a) proposes a direct approach where the object of modeling is the conditional expectation of a certain nonlinear function of the variable being forecast. Lastly, a long-debated issue in time series multiperiod forecasting literature is whether iterated or direct forecasts are better (e.g., Marcellino, Stock and Watson, 2006). Here, the object of interest is the multiperiod conditional expectation, and modeling it constitutes the direct

[1] This approach lies within the "decisionmetrics" paradigm of Skouras (2007), where an econometric model is developed so that it serves a particular purpose rather than is used in a variety of contexts.

approach, while a semi-direct approach assumes modeling the one-period ahead conditional expectation and iteratively deducing the multiperiod conditional expectation.

In this paper we investigate, using experiments with real data, the question of whether the return signs, or equivalently, directions-of-change, are better to forecast using a direct, semi-direct, or indirect approach. Direction-of-change forecasts are useful in formation if trading strategies and efficient asset allocation and have lately received a lot of attention of financial econometricians (e.g., Rydberg and Shephard, 2003; Pesaran and Timmermann, 2004; Christoffersen and Diebold, 2006; Chung and Hong, 2007; Anatolyev, 2009b; Anatolyev and Gospodinov, 2010). First, directional forecasts can be generated indirectly by reading off the conditional distribution, see Christoffersen, Diebold, Mariano, Tay and Tse (2007) and Bekiros and Georgoutsos (2008), among others. Second, they can be produced by a dynamic model for certain conditional objects, more complex than the up/down conditional probability, and this would constitute the semi-direct approach. Finally, the conditional probabilities can be modeled directly in a binary autoregressive framework, see Startz (2008).

Importantly, we perform the comparisons assuming that the underlying loss function is asymmetric, corresponding to the so called Linlin loss. While the symmetric quadratic (Quad) loss function $\frac{1}{2}u^2$ is still prevailing in econometrics because of its convenience and tractability, more and more often researchers use asymmetric loss functions in empirical analysis, such as the linear-exponential (Linex) loss of the form $\exp(\theta u) - \theta u - 1$ or the doubly linear (Linlin) loss of the form $\left((1-\alpha)\,\mathbb{I}_{\{u<0\}} + \alpha\mathbb{I}_{\{u>0\}}\right)|u|$ (where the parameters θ or α index the degree of asymmetry) more adequately reflecting asymmetries in preferences of decision makers. Empirical plausibility is the first reason of our use of the Linlin loss. The second reason is the accumulated experience due to the increased interest to modeling and analyzing the quantiles of the conditional distribution of returns, see for example Engle and Manganelli (2004), Lee and Yang (2006) and Cenesizoglu and Timmermann (2008). An important, albeit technical, reason why we prefer Linlin to other asymmetric losses is that a quantile predictor is directly linked to a direction-of-change predictor (Granger and Pesaran, 2000; Lee and Yang, 2006). Note that our interest is to those quantiles that are not too far from the median, which significantly differs from that in the VaR analysis where the focus is on quantiles in the tails.

We do real data experiments with the S&P500 index and DM/USD exchange rate of various frequencies: weekly, daily and intradaily. As the "indirect" model for the predictive density, we use the flexible NGARCHSK class of León, Rubio and Serna (2005) (see also León, Mencía, and Sentana, 2009). As the "semi-direct" model for the quantiles, we use the CAViAR class of Engle and Manganelli (2004). Finally, as the "direct" model for directional indicators, we use the BARMA class of Startz (2009). By and large, the semi-direct approach, i.e. by way of modeling the evolution of conditional quantiles, turns out to be markedly superior to the direct and indirect approaches for stock returns, especially at daily frequency, and not worse for exchange rate returns. We also run additional experiments in order to see how robust these tendencies are, where we exploit variations of the same models that deviate from the baseline ones in minor ways.

The paper is structured as follows. Section 2 describes the theory related to the relationship among the three approaches to forecasting the directions-of-change. In Section 3 the corresponding models are described. Section 4 contains the description of data and the results. Finally, Section 5 concludes.

2 Directional forecasting under asymmetric loss

2.1 Link between directional and return forecasts

Let $\{r_t\}_{t=1}^T$ be the series of financial returns. Also we consider the binary return indicator, or direction-of-change, series $\{y_t\}_{t=1}^T$,

$$y_t = \mathbb{I}\{r_t \geq 0\} = \frac{\text{sgn}(r_t) + 1}{2},$$

where $\mathbb{I}\{A\}$ is the indicator of event A equalling one if A is true and zero otherwise, and $\text{sgn}(u)$ is a sign function equalling 1 if u is non-negative and -1 otherwise.

Consider a forecaster who makes directional and return forecasts for the same return series. Let $\widehat{r}_{t+1|t}$ be an optimal forecast of the return r_{t+1} at $t+1$ made at t, while $\widehat{y}_{t+1|t}$ be her optimal directional forecast, i.e. of the indicator y_{t+1} at $t+1$ made at t. Of course, the forecaster predicts the market to move up when the return forecast is positive, and to move down otherwise. Thus, the optimal directional and return forecasts are linked in the following way:

$$\widehat{y}_{t+1|t} = \mathbb{I}\left\{\widehat{r}_{t+1|t} \geq 0\right\}. \tag{1}$$

The relation (1) allows us to employ the semi-direct approach: when there is a model for return levels that generates a return forecast, this return forecast can be translated into the directional forecast using (1).

2.2 "Continuous" and "discrete" losses

Let the forecaster be endowed with the "continuous" loss function $c(r_{t+1} - \widehat{r}_{t+1|t})$ when she evaluates return forecasts. The forecast $\widehat{r}_{t+1|t}$ introduced above is optimal in the sense of minimizing this "continuous" loss. Note that traditionally the only argument is the difference between the return realization and its forecast. When the forecaster makes directional forecasts, she is implicitly driven by some underlying "discrete" loss function $d_t(\widehat{y}_{t+1|t}, y_{t+1})$, or, equivalently, "discrete" utility function $u_t(\widehat{y}_{t+1|t}, y_{t+1})$, and $\widehat{y}_{t+1|t}$ is optimal in the sense of minimizing this "discrete" loss or maximizing this "discrete" utility. In contrast to the "continuous" counterpart, in general these "discrete" functions are also functions of the information set Ω_t available to the forecaster at t, the period of making the forecast (hence the index $_t$), and the argument is not necessarily single.

The most general "discrete" utility function has the form (Granger and Pesaran, 2000; Elliott and Lieli, 2005)

$$u_t(\widehat{y}_{t+1|t}, y_{t+1}) = \begin{cases} u_{1,+,t} & \text{if } \widehat{y}_{t+1|t} = y_{t+1} = 1, \\ u_{0,-,t} & \text{if } \widehat{y}_{t+1|t} = y_{t+1} = 0, \\ u_{1,-,t} & \text{if } \widehat{y}_{t+1|t} = 1,\ y_{t+1} = 0, \\ u_{0,+,t} & \text{if } \widehat{y}_{t+1|t} = 0,\ y_{t+1} = 1, \end{cases}$$

which can be alternatively represented in the form of the following 2×2 payoff matrix:

$u_t(\widehat{y}_{t+1|t}, y_{t+1})$	$r_{t+1} \geq 0$	$r_{t+1} < 0$
$\widehat{y}_{t+1|t} = 1$	$u_{1,+,t}$	$u_{1,-,t}$
$\widehat{y}_{t+1|t} = 0$	$u_{0,+,t}$	$u_{0,-,t}$

2.3 Link between directional and probability forecasts

Let us define the conditional probability of an up movement as

$$\pi_t = \Pr\{r_{t+1} \geq 0 | \Omega_t\},$$

where Ω_t is the information set containing r_t and its past. Note that

$$\pi_t = E(y_{t+1}|\Omega_t).$$

In order to employ the both direct and indirect approaches, we need to tie the directional forecasts to this conditional probability.

The expected utility of the decision $\widehat{y}_{t+1|t} = 1$ is $u_t(1, \cdot) = (1 - \pi_t)u_{1,-,t} + \pi_t u_{1,+,t}$, while that of the decision $\widehat{y}_{t+1|t} = 0$ is $u_t(0, \cdot) = (1 - \pi_t)u_{0,-,t} + \pi_t u_{0,+,t}$. Then the forecaster will prefer the forecast $\widehat{y}_{t+1|t} = 1$ if $u_t(1, \cdot) \geq u_t(0, \cdot)$ which is equivalent to the rule

$$\pi_t \geq \frac{u_{0,-,t} - u_{1,-,t}}{(u_{0,-,t} - u_{1,-,t}) + (u_{1,+,t} - u_{0,+,t})} \equiv \bar{\pi}_t.$$

Thus the optimal indicator forecast is

$$\widehat{y}_{t+1|t} = \mathbb{I}\{\pi_t \geq \bar{\pi}_t\}. \tag{2}$$

The (generally time varying) threshold $\bar{\pi}_t$ is completely determined by the "discrete" utility function in hand. The relation (2) allows us to employ the direct and indirect approaches: when there is a model for the conditional probability or conditional density that generates a probability forecast, this probability forecast can be translated into the directional forecast using (2). The comparison of (1) and (2) gives the conclusion that the events $\widehat{r}_{t+1|t} \geq 0$ and $\pi_t \geq \bar{\pi}_t$ are equivalent for an optimizing forecaster.

Furthermore, Granger and Pesaran (2000) show (see also Lee and Yang, 2006) that the optimal directional predictor minimizes the expected "discrete" loss $E[d_t(v_{t+1})]$, where $v_{t+1} = y_{t+1} - \widehat{y}_{t+1|t}$ is an indicator forecast error, and $d_t(v)$ is a "discrete" loss function

$$d_t(v) = \begin{cases} 1 - \bar{\pi}_t & \text{if } v = 1, \\ \bar{\pi}_t & \text{if } v = -1, \\ 0 & \text{if } v = 0, \end{cases}$$

or in the 2×2 payoffs matrix form,

d_t	$y_{t+1} = 1$	$y_{t+1} = 0$
$\widehat{y}_{t+1\|t} = 1$	0	$\bar{\pi}_t$
$\widehat{y}_{t+1\|t} = 0$	$1 - \bar{\pi}_t$	0

This "discrete" loss function will allow us to evaluate directional forecasts.

2.4 Asymmetric loss

The most widespread loss function is quadratic: $c(u) = \frac{1}{2}u^2$. Its popularity can be explained by a simple form of optimal predictor, the conditional mean of returns, and tractability due to the linearity of the first derivative. However, in real life symmetric loss functions do not correspond to the actual behavior of economic agents. Examples of such situations are given by Granger (1969), Capistrán-Carmona (2005), Elliott, Komunjer and Timmermann (2008), and others.

Among asymmetric loss functions are linear exponential (Linex) $c(u) = \exp(\theta u) - \theta u - 1$, $\theta \neq 0$ and doubly quadratic (Quadquad) $c(u) = \left((1-\varphi)\mathbb{I}_{\{u<0\}} + \varphi \mathbb{I}_{\{u>0\}}\right) u^2$, $\varphi \in (0,1)$ as well as the most popular doubly linear (Linlin) loss

$$c(u) = \left((1-\alpha)\mathbb{I}_{\{u<0\}} + \alpha \mathbb{I}_{\{u>0\}}\right) |u|, \quad \alpha \in (0,1). \tag{3}$$

In all these cases an additional known parameter θ, φ or α is present that indicates the degree of asymmetry. The Linex and Quadquad loss functions are not robust to outliers, especially the Linex one because of the presence of exponent. The optimal predictor under Linex is quite involved and is a certain nonlinear transformation of the conditional expectation of exponent of the variable being forecast (Zellner, 1986), and moment requirements may not hold when the Linlin loss applied to financial data. At the same time, while the Quadquad loss is less prone to the effects of heavy tails, the closed-form optimal predictor does not exist in this case (Christoffersen and Diebold, 1996).

The "tick" function corresponding to Linlin does not have these shortcomings. The corresponding optimal predictor is a conditional α-quantile $q_\alpha(r_{t+1}|\Omega_t)$ which is a much more robust regression measure. In addition, the fact that the quantiles are important in the VaR analysis, makes the Linlin loss very popular and important. The drawback of the Linlin loss functions is non-differentiability at zero which complicates estimation and inference to a certain degree. However, recent financial econometric literature has been showing an increased interest in modeling conditional quantiles (e.g., McNeil and Frey, 2000; Engle and Manganelli, 2004; Kuester, Mittnik and Paolella, 2006).

There is also another, more technical reason why we prefer the Linlin loss function. The threshold $\bar{\pi}_t$ that links the conditional "success" probability to the optimal sign forecast via (2) is generally time varying and may have various forms. For example, for the commonly used quadratic loss, it equals

$$-\frac{E(r_{t+1}|r_{t+1}<0,\Omega_t)}{E(r_{t+1}|r_{t+1}\geq 0,\Omega_t)-E(r_{t+1}|r_{t+1}<0,\Omega_t)},$$

while for the Linex loss it is

$$\frac{1-E(e^{\alpha r_{t+1}}|r_{t+1}<0,\Omega_t)}{E(e^{\alpha r_{t+1}}|r_{t+1}\geq 0,\Omega_t)-E(e^{\alpha r_{t+1}}|r_{t+1}<0,\Omega_t)}.$$

In these two examples, to compute $\bar{\pi}_t$ one additionally needs models for certain complicated conditional expectations. In contrast, for the Linlin loss the threshold $\bar{\pi}_t$ is simply

$$1-\alpha,$$

which is not only time invariant, but also known in advance. This is established, in particular, in Lee and Yang (2006).

3 Forecasting models and forecasts

We use three types of models with decreasing complexity of the modeled object in order to produce forecasts of return indicators. The most complex model (the "density model", D) describes the dynamics of the conditional density of returns. The less complex model (the "return model", R) describes the dynamics of return levels. Finally, the simplest model (the "sign model", S) describes the dynamics of return signs themselves. Put differently, the D model produces indirect sign forecasts, the R model produces semi-direct sign forecasts, and the S model produces direct sign forecasts.

3.1 The "density model"

As the D model we use the NGARCHSK model of León, Rubio and Serna (2005):

$$r_t = \mu_t + \varepsilon_t, \quad \varepsilon_t = h_t^{1/2}\eta_t, \quad \eta_t|\Omega_{t-1} \sim \mathcal{F}_t,$$

where the conditional distribution $\mathcal{F}_t$ is described in terms of the conditional density as follows:

$$f_t(x) = \frac{\varphi(x)\psi_t^2(x)}{\Gamma_t},$$

where $\varphi(x)$ is the standard normal density,

$$\psi_t(x) = 1 + \frac{s_t}{3!}(x^3 - 3x) + \frac{k_t - 3}{4!}(x^4 - 6x^2 + 3)$$

comes from the Gram–Charlier expansion,

$$\Gamma_t = 1 + \frac{s_t^2}{3!} + \frac{(k_t - 3)^2}{4!}$$

is a normalizing term, and s_t and k_t are associated with the conditional third and fourth order moments. The NGARCHSK model is a flexible fully parametric model in the spirit of Hansen's (1994) autoregressive conditional density (ARCD), which models not only the conditional mean and variance, but also time-varying skewness and kurtosis. The idea behind it is to use the Gallant and Tauchen (1989) seminonparametric family of densities as a parametric class (see also León, Mencía, and Sentana, 2009). These densities are based on the Gram–Charlier expansion around the normal density. After squaring and renormalization the resulting conditional density $f_t(x)$ is automatically a valid density function, as it is non-negative and integrates to one. One can construct the loglikelihood function in a straightforward way. Another advantage is that this class nests the normal density corresponding to the case $s_t = 0$ and $k_t = 3$.

For the conditional mean we use a linear AR(1) specification which is traditionally employed for financial data in order to capture slight autocorrelatedness:

$$\mu_t = \mu + \rho r_{t-1}.$$

The conditional second moment follows a GARCH-type dynamics as in León, Rubio and Serna (2005):

$$h_t = \beta_0 + \beta_1 h_{t-1} + \beta_2(\varepsilon_{t-1} + \beta_3\sqrt{h_{t-1}})^2,$$

hence the familiar GARCH letters in the acronym NGARCHSK. Next, the letter N stands for "nonlinear". The nonlinear term in the variance equation accounts for the leverage effect. Its form is taken from Engle and Ng (1993), and the coefficient β_3 turns out to be statistically significant for all financial series used in León, Rubio and Serna (2005). Finally, the letter S in the acronym NGARCHSK stands for "skewness" and K for "kurtosis" which indicates the possibility that these conditional characteristic are allowed to be time varying. In our basic density specification, however, while we keep the nonlinear dynamics for the conditional variance, we set the conditional third and fourth moments to be constant: $s_t = \sigma_0$, $k_t = \kappa_0$. This is because the empirical performance with these moments constant is not worse than with varying ones. We however also try time varying conditional third and fourth moments when we check for robustness (see subsection 4.4).

The NGARCHSK model is, of course, not the only way to flexibly parameterize the conditional density. Earlier Harvey and Siddique (1999) proposed a flexible parameterization on the basis of non-central t distribution which allowed for time varying conditional skewness but not conditional kurtosis. Other parameterizations with time-varying skewness and kurtosis have been suggested in the literature as well, for example, in Jensen and Lunde (2001) and Wilhelmsson (2009). The NGARCHSK model, however, has a more intuitive design and bigger flexibility (see León, Mencía, and Sentana, 2009).

The directional forecast $\widehat{y}_{t+1|t}^{(D)}$ is extracted from $\widehat{f}_t(x)$, the estimated conditional density of standardized errors η_t, as follows:

$$\widehat{y}_{t+1|t}^{(D)} = \mathbb{I}\{\widehat{\pi}_t^{(D)} \geq 1 - \alpha\},$$

where the predictor $\widehat{\pi}_t^{(D)}$ of the positive return probability π_t is obtained from the estimated predictive density via integration:

$$\widehat{\pi}_t^{(D)} = \int_0^{+\infty} \widehat{g}_t(r)dr,$$

where $\widehat{g}_t(x)$ is the estimated predictive density of r_{t+1}, which is a transformation of $\widehat{f}_t(x)$. The integration is performed numerically.

3.2 The "return model"

As the R model we utilize the Conditional Autoregressive Value at Risk (CAViAR) model of Engle and Manganelli (2004) for conditional quantiles $q_{\alpha,t} \equiv q_\alpha(r_t|\Omega_{t-1})$. The general CAViAR$(p,s)$ specification has the GARCH-type autoregressive dynamics

$$q_{\alpha,t} = \lambda + \sum_{i=1}^{p} \psi_i q_{\alpha,t-i} + \sum_{i=1}^{s} \chi_i l_{t-i},$$

where the driving process l_t is a function of a finite number of observations from Ω_t. Engle and Manganelli (2004) suggest several specifications of the driving process tied to the VaR nature of the variables, i.e. very small α like 1% or 5%. In contrast, we are interested in a central tendency, i.e. in middle sized α. In this light, we select the most reasonable, called asymmetric slope, version of the CAViAR$(1,1)$ model

$$q_{\alpha,t} = \gamma_0 + \gamma_1 q_{\alpha,t-1} + \gamma_2 r_{t-1}^{+} + \gamma_3 r_{t-1}^{-},$$

where $x^+ = \max(x,0)$ and $x^- = -\min(x,0)$. Engle and Manganelli (2004) discuss other three specifications: adaptive slope, symmetric absolute value, indirect GARCH. The asymmetric slope version is more general than the symmetric absolute value model and reflects the leverage effect. We also try the adaptive slope version when we check for robustness (see subsection 4.4).

A sign forecast $\widehat{y}_{t+1|t}^{(R)}$ is generated from the estimated quantile prediction $\widehat{q}_{\alpha,t+1|t}$ by the familiar rule:

$$\widehat{y}_{t+1|t}^{(R)} = \mathbb{I}\{\widehat{q}_{\alpha,t+1|t} \geq 0\}.$$

3.3 The "sign model"

As the S model we use the binary autoregressive moving average (BARMA) model of Startz (2008) where the directions-of-change are modeled directly as functions of their own past:

$$\pi_t = \frac{\exp(\theta_t)}{1+\exp(\theta_t)}, \qquad \theta_t = \lambda + \sum_{i=1}^{p} \psi_i y_{t-i} + \sum_{i=1}^{q} \chi_i \theta_{t-i}.$$

We set the orders p and q to unity. The logit link is traditional in financial applications (e.g., Rydberg and Shephard, 2003; Christoffersen and Diebold, 2006; Anatolyev and Gospodinov, 2010), while the past indicators usually perform better as driving variables than, say, past returns.

The BARMA model generates probability forecasts $\pi_t^{(S)}$ whose estimated values $\widehat{\pi}_t^{(S)}$ are used to produce the sign forecasts by the same rule as in the D model:

$$\widehat{y}_{t+1|t}^{(S)} = \mathbb{I}\{\widehat{\pi}_t^{(S)} \geq 1 - \alpha\}.$$

4 Empirical evidence

4.1 Data

We use the following six time series, each having $T = 1200$ observations.

- Returns S&P500 (SP500m). Frequency: 15 minutes. Time period: from 01.02.2007 to 11.04.2007. Source: `finam.ru`.
- Exchange rates DM to USD (DMUSDh). Frequency: hourly. Time period: from 22.03.2007 to 11.06.2007. Source: `finam.ru`.
- Returns S&P500 (SP500d). Frequency: daily. Time period: from 26.01.1989 to 22.10.1993. Source: `finance.yahoo.com`.
- Exchange rates DM to USD (DMUSDd). Frequency: daily. Time period: from 02.01.1990 to 15.04.1993. Source: `oanda.com`.
- Returns S&P500 (SP500w). Frequency: weekly. Time period: from 03.01.1950 to 02.01.1973. Source: `finance.yahoo.com`.
- Exchange rates DM to USD (DMUSDw). Frequency: weekly. Time period: from 19.12.1984 to 26.12.2007. Source: `oanda.com`.

The start and end dates are picked without reference to any particular reasons, with an eye only on data availability. Here we have two series of ultra high frequency, two series of daily returns and two series of weekly returns. Figure 1 depicts the return series on the left side. The right side of Figure 1 shows cumulative sign series (i.e. values of the sign function accumulated from the start to the present dates). Descriptive statistics are presented in the following table. All return series show unconditional leptokurtosis and skewness of various degree, tending to be higher for higher frequencies. The stock market was largely bullish, while the exchange market went up and down exhibiting slight appreciation of the dollar in the long run specific for each frequency.

	SP500m	DMUSDh	SP500d	DMUSDd	SP500w	DMUSDw
mean	0.00	−0.00	0.04	0.00	0.16	0.07
median	0.00	0.00	0.03	0.00	0.28	0.06
standard deviation	0.11	0.05	0.58	0.61	1.28	1.23
skewness	1.36	−0.33	−0.18	−0.36	−0.51	0.07
kurtosis	40.94	17.14	5.48	6.13	4.10	4.25
how many up	611	605	630	610	718	623
how many down	576	586	569	588	481	576

From the return statistics and graphs one can see that using conditional quantiles (and hence implicitly the Linlin loss), the conditional median in particular, may be preferred to using conditional moments, the conditional mean in particular, as the former exhibit clear robustness and always exist while the latter may not. These issues may be quite important especially for the highest frequency data where outliers are quite pronounced.

4.2 Forecasting procedure

We use the first $R = 1000$ observations for in-sample modeling and the rest $P = 200$ observations for an out-of-sample forecasting experiment. We use the rolling scheme of generating out-of-sample forecasts. That is, when the p^{th} forecast is made, $p = 1, ..., P$, the estimates are recomputed using observations from $t = p$ to $t = R + p - 1$. All programs are written and run in MATLAB.[2]

We do all experiments for α from the following grid: 0.30, 0.40, 0.45, 0.50, 0.55, 0.60, 0.70. We pay more attention to α near 0.50 and hence quantiles near the median. Larger deviations from 0.50 are less interesting as they are empirically less plausible, and for α too

[2]For CAViAR models, we have used Simone Manganelli's code.

high or too low there are too little observations on one of sides of the distribution rendering statistics collected from such samples unreliable. Note that Cenesizoglu and Timmermann (2008) document larger predictability of return quantiles, albeit for monthly stock returns, for larger deviations of α from 0.50.

The quality criterion is the average value of the "discrete" loss function

$$d(v) = \begin{cases} \alpha & \text{if } v = 1, \\ 1-\alpha & \text{if } v = -1, \\ 0 & \text{if } v = 0, \end{cases}$$

applied to the forecast errors. Thus, the in-sample criterion is

$$DL_{in} = \frac{1}{R}\sum_{t=1}^{R} d(y_t - \widehat{y}_{t|t-1}),$$

where $R = 1,000$ is the size of the (initial) estimation subsample. The out-of-sample criterion is

$$DL_{out} = \frac{1}{P}\sum_{t=T-P+1}^{T} d(y_t - \widehat{y}_{t|t-1}),$$

where $P = 200$ is the size of the forecasting subsample. The letters d and D above stand for "discrete". Smaller values of DL imply better performance. It is this loss function that is consistent with the optimizing behavior of agents, and it would be ridiculous to use other performance measures (for example, the one that just counts "successes" and "failures").

As simplest benchmarks we use trivial directional forecasts: one that always predicts 0 (i.e. down) and one that always predicts 1 (i.e. up).

4.3 Empirical results

Table 1 presents criteria values attained for different configurations. The information is arranged in the following way. Part (a) refers to the highest frequency data, part (b) to the daily data, and part (c) to the weekly data. In each, the upper part relates to the SP500 index, the lower part – to the DM/USD exchange rate; the left half – to the in-sample computations, the right half – to the out-of-sample computations. The minimal criterion value(s) across each half of each line is in boldface. Figure 2 shows the cumulative "discrete" loss (i.e. values of the loss function accumulated from the first to the present forecast dates) for out-of-sample sign forecasts of the SP500 index returns, the most clear-cut case, for selected values of α: two on the opposite sides of the conditional distribution and one implying exactly the conditional median. Several important observations follow.

For the stock market returns characterized by some perceptible predictability the semi-direct approach provides much better directional forecasts. The superiority of the semi-direct approach is much sharper for the daily frequency than for the other two frequencies. The ranking of the other two approaches, indirect and direct, is fuzzy, although by and large the sign model tends to produce directional forecasts of a bit better quality. The superiority of the semi-direct approach is more pronounced for less extreme values of α. This evidence is a bit unexpected in light of Cenesizoglu and Timmermann (2008) who discover less predictability of conditional quantiles near the conditional median. However, Cenesizoglu and Timmermann (2008) used monthly data and prediction by exogenous predictors; in addition, indirect and direct approach may be prone to the same tendency even more.

For the exchange rate returns characterized by little, if any, predictability, the differences across approaches are much more blurry. Most blurry they are for the daily frequency, while in the case of higher frequency the semi-direct approach is a little better, at least in-sample, and in the case of weekly frequency the sign forecasts are a bit better, at least out-of-sample. The differences though, if any, are small in magnitude. The differences across approaches are also most blurry for higher deviations of α from 0.50, when sometimes trivial forecasts more often are not worse than model-based forecasts.

In it important to note that, in general, the discovered patterns agree in in-sample and out-of-sample experiments. While the customary tension between in-sample and out-of-sample predictability may make one expect disagreement, this tension evidently does not apply to the comparison across *approaches*.

4.4 Robustness check

We run some experiments with model modifications to make sure that our numbers are robust to minor deviations in specifications, so that adding or removing some parametric elements do not change our conclusions dramatically. We restrict ourselves to experimentation with the daily S&P500 index. The results are shown in Table 2.

First we try to remove the AR(1) component from the conditional mean, or replace it with an ARCH-M term δh_t. Note that the former modification is equivalent to a random walk with a complex density superimposed on the innovations. Both modifications lead to significantly worse forecast performance (not shown), both in and out of sample, and the forecasts practically coincide with trivial ones. It seems that there is some dynamics in the conditional mean, but it can be successfully captured by the linear autoregressive term.

In columns 2 to 4 of Table 2 we give the results for various density models. Column 2 replicates the results for the basic NGARCHSK specification from Table 1b, with constant skewness and kurtosis parameters. Column 3 contains those when the skewness and kurtosis are allowed to be time-varying as in León, Rubio and Serna (2005):

$$\begin{aligned} s_t &= \sigma_0 + \sigma_1 s_{t-1} + \sigma_2 \eta_{t-1}^3, \\ k_t &= \kappa_0 + \kappa_1 k_{t-1} + \kappa_2 \eta_{t-1}^4. \end{aligned}$$

It is worth noting that the parameters $\sigma_1, \sigma_2, \kappa_1, \kappa_2$ are statistically insignificant in roughly half of cases in León, Rubio and Serna (2005) when the authors apply the model to returns from stock indices and exchange rates but no particular pattern emerges. The resulting differences in forecast performance, both in and out of sample, are different in direction but small in magnitude. Hence, time variation in higher order conditional moment is not that significant to justify estimation of four extra parameters.

Next we exploit the simple normal AR(1)–GARCH(1,1) model whose variance equation is plainly

$$h_t = \beta_0 + \beta_1 h_{t-1} + \beta_2 \varepsilon_{t-1}^2.$$

This is a most simple volatility model which is often used in practice. The results are shown in Column 4. The values of forecast performance criteria are a bit larger than in the previous case; occasional minor improvements however fall short of the solid improvements provided by the semi-direct approach.

As for the return model, we have also tried the adaptive slope variation of the CAViaR model also used in Engle and Manganelli (2004):

$$q_{\alpha,t} = q_{\alpha,t-1} + \gamma_1 \left(\frac{1}{1 + \exp\left(G\left(r_{t-1} - q_{\alpha,t-1}\right)\right)} - \alpha \right),$$

where $G = 10$. This specification is tied to small α, and the corresponding forecasts practically coincide with trivial ones, hence not shown. It is clear that careful specification of the dynamics of conditional quantiles is a key to its superior performance.

Finally, we check the robustness to an order specification of BARMA models, see Columns 5 and 6 in Table 2. While Column 5 replicates the results of fitting the BARMA(1,1) equation from Table 1b, Column 6 contains those for a higher order sign model, BARMA(2,2). Allowing higher orders does not practically change the forecasts.

Thus, our earlier conclusion of the superiority of the semi-direct approach for stock returns remains valid. As for the exchange rate returns, although we did not run similar experiments with the DM/USD series, it is clear that changing dynamic specifications even in minor ways is able to kill or revert those tiny discrepancies between the approaches when predictability is low.

5 Concluding remarks

Modeling and estimating the entire conditional density (constituting the indirect approach), with the noise arising from modeling uncertainty and estimation error, turns out to be too complex task for prediction of directions-of-change. On the other hand, existing models for binary up/down indicators (constituting the direct approach) do not possess flexibility sufficient to generate reliable forecasts. An intermediate way, by modeling and estimating the dynamics of conditional quantiles (constituting the semi-direct approach), proves to be more effective as these models warrant an optimal degree of flexibility and parsimony. This is especially true for stock returns as a relatively more predictable series, while the patterns for exchange rate returns are much more fuzzy.

Acknowledgments

We thank Nikolay Gospodinov for useful discussions, and Simone Manganelli for making his CAViAR codes publicly available.

References

[1] Anatolyev, S. (2009a) Dynamic modeling under linear-exponential loss. *Economic Modeling* 26, 82–89.

[2] Anatolyev, S (2009b) Multi-market direction-of-change modeling using dependence ratios. *Studies in Nonlinear Dynamics & Econometrics*, Vol. 13, issue 1, article 5.

[3] Anatolyev, S. and N. Gospodinov (2010) Modeling financial return dynamics via decomposition. *Journal of Business & Economic Statistics*, forthcoming.

[4] Bekiros, S. and D.A. Georgoutsos (2008) Direction-of-change forecasting using a volatility based recurrent neural network. *Journal of Forecasting* 27, 407–417.

[5] Capistrán-Carmona, C. (2005) A Review of Forecasting Theory using Generalized Loss Functions. Working paper, University of California, San Diego.

[6] Christoffersen, P.F. and Diebold, F.X. (1996) Further Results on Forecasting and Model Selection Under Asymmetric Loss. *Journal of Applied Econometrics* 11, 561–571.

[7] Christoffersen, P.F. and Diebold, F.X. (2006) Financial Asset Returns, Direction-of-Change Forecasting, and Volatility Dynamics. *Management Science* 52, 1273–1287.

[8] Christoffersen, P.F., Diebold, F.X., Mariano, R.S., Tay, A.S. and Tse, Y.K. (2007) Direction-of-Change Forecasts Based on Conditional Variance, Skewness and Kurtosis Dynamics: International Evidence. *Journal of Financial Forecasting* 1, 1–22.

[9] Cenesizoglu, T. and A. Timmermann (2008) Is the Distribution of Stock Returns Predictable? Manuscript, UCSD.

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

[11] Elliott, G. and Lieli, R.P. (2005) Predicting Binary Outcomes. Working paper, University of California, San Diego.

[12] Elliott, G., Komunjer, I. and Timmermann, A. (2008) Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss? *Journal of the European Economic Association* 6, 122–157.

[13] Engle, R.F. and Ng, V.K. (1993) Measuring and testing the impact of news on volatility. *Journal of Finance* 48, 1749–1778.

[14] Engle, R.F., Manganelli, S. (2004) CAViAR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business and Economic Statistics* 22, 367–381.

[15] Gallant, A.R. and Tauchen, G. (1989). Seminonparametric estimation of conditionally constrained heterogeneous processes: asset pricing applications. *Econometrica* 57, 1091–1120.

[16] Granger, C.W.J. (1969) Prediction with a Generalized Cost of Error Function. *Operational Research Quarterly* 20, 199–207.

[17] Granger, C.W.J. and Pesaran, M.H. (2000) Economic and Statistical Measures of Forecast Accuracy. *Journal of Forecasting* 19, 537–560.

[18] Hansen, B. (1994) Autorregressive conditional density estimation. *International Economic Review* 35, 705–730.

[19] Harvey, C.R. and Siddique, A. (1999) Autoregressive Conditional Skewness. *Journal of Financial and Quantitative Analysis* 34, 465–487.

[20] Jensen, M.B. and A. Lunde (2001) The NIG-S&ARCH model: A fat tailed, stochastic, and autoregressive conditional heteroskedastic volatility model. *Econometrics Journal* 4, 319–342.

[21] Kuester, K., S. Mittnik, and M.S. Paolella (2006) Value-at-risk prediction: a comparison of alternative strategies. *Journal of Financial Econometrics*, 4, 53–89.

[22] Lee, T.-H. and Yang, Y. (2006) Bagging Binary and Quantile Predictors for Time Series. *Journal of Econometrics* 135, 465–497.

[23] León, Á., Rubio, G. and Serna, G. (2005) Autoregresive Conditional Volatility, Skewness and Kurtosis. *The Quarterly Review of Economics and Finance* 45, 599–618

[24] León, Á., Mencía, J. and Sentana, E. (2009) Parametric Properties of Semi-Nonparametric Distributions, with Applications to Option Valuation, *Journal of Business and Economic Statistics* 27, 176–192.

[25] Marcellino, M., Stock, J.H. and M. Watson (2006) A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics* 135, 499–526.

[26] McNeil, A. J. and R. Frey (2000). Estimation of Tail-related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach. *Journal of Empirical Finance* 7, 271–300.

[27] Patton, A.J., and A. Timmermann (2007) Properties of optimal forecasts under asymmetric loss and nonlinearity. *Journal of Econometrics*, 140, 884–918.

[28] Pesaran, M.H. and A. Timmermann (2004) How costly is it to ignore breaks when forecasting the direction of a time series? *International Journal of Forecasting* 20 411–425.

[29] Rydberg, T.H., and N. Shephard (2003) Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics* 1, 2–25.

[30] Skouras, S.P. (2007) Decisionmetrics: a decision-based approach to econometric modelling. *Journal of Econometrics*, 137, 414–440.

[31] Startz, R. (2008) Binomial Autoregressive Moving Average Models with an Application to U.S. Recessions. *Journal of Business* & *Economic Statistics* 26, 1–8.

[32] Timmermann, A. (2008) Elusive return predictability. *International Journal of Forecasting* 24, 1–18.

[33] Wilhelmsson, A. (2009) Value at Risk with time varying variance, skewness and kurtosis–the NIG-ACD model. *Econometrics Journal* 12, 82–104

[34] Zellner, A. (1986) Bayesian Estimation and Prediction Under Asymmetric Loss Function. *Journal of American Statistical Association* 81, 446–451.

Table 1a. In-sample and out-of-sample, high frequency data

α	In-sample loss					Out-of-sample loss				
	Dynamic models			Trivial forecast		Dynamic models			Trivial forecast	
	D	R	S	1	0	D	R	S	1	0
					SP500m					
0.30	0.154	0.152	**0.151**	0.347	**0.151**	**0.150**	0.178	0.177	0.280	0.180
0.40	0.180	**0.179**	0.189	0.298	0.202	0.194	**0.179**	0.190	0.240	0.240
0.45	0.188	**0.179**	0.189	0.273	0.227	0.210	**0.179**	0.186	0.220	0.270
0.50	0.189	**0.174**	0.190	0.248	0.252	0.188	**0.168**	0.185	0.200	0.300
0.55	0.182	**0.172**	0.192	0.223	0.277	**0.170**	0.172	0.194	0.180	0.330
0.60	0.172	**0.168**	0.187	0.198	0.302	0.166	**0.147**	0.158	0.160	0.360
0.70	0.145	**0.129**	0.149	0.149	0.353	0.125	0.121	**0.119**	0.120	0.420
					DMUSDh					
0.30	0.149	**0.147**	0.154	0.340	0.154	**0.150**	**0.150**	**0.150**	0.350	**0.150**
0.40	0.184	**0.179**	0.200	0.292	0.206	0.190	**0.175**	0.215	0.300	0.200
0.45	0.192	**0.185**	0.205	0.267	0.231	**0.193**	0.197	0.218	0.275	0.225
0.50	0.203	**0.187**	0.205	0.243	0.257	**0.195**	0.200	0.230	0.250	0.250
0.55	0.198	**0.186**	0.206	0.219	0.283	0.195	**0.183**	0.246	0.225	0.275
0.60	0.185	**0.176**	0.179	0.194	0.308	0.183	**0.170**	0.271	0.200	0.300
0.70	0.151	**0.142**	0.146	0.146	0.360	**0.146**	0.152	0.275	0.150	0.350

Notes: Table shows discrete loss values. The upper part relates to the SP500 index, the lower part – to the DM/USD exchange rate; the left half – to the in-sample computations, the right half – to the out-of-sample computations. The minimal criterion value(s) across each half of each line is in boldface.

Table 1b. In-sample and out-of-sample, daily data

α	In-sample loss					Out-of-sample loss				
	Dynamic models			Trivial forecast		Dynamic models			Trivial forecast	
	D	R	S	1	0	D	R	S	1	0
					SP500d					
0.30	0.138	**0.105**	0.144	0.335	0.157	0.144	**0.109**	0.145	0.319	0.164
0.40	0.150	**0.120**	0.165	0.287	0.209	0.174	**0.114**	0.170	0.273	0.218
0.45	0.154	**0.121**	0.164	0.263	0.235	0.194	**0.119**	0.175	0.250	0.245
0.50	0.166	**0.123**	0.164	0.239	0.261	0.173	**0.120**	0.175	0.228	0.273
0.55	0.166	**0.124**	0.160	0.215	0.287	0.166	**0.129**	0.171	0.205	0.300
0.60	0.160	**0.121**	0.163	0.191	0.313	0.159	**0.129**	0.167	0.182	0.327
0.70	0.135	**0.109**	0.129	0.143	0.365	0.126	**0.113**	0.129	0.137	0.382
					DMUSDd					
0.30	**0.151**	**0.151**	**0.151**	0.347	**0.151**	**0.162**	**0.162**	**0.162**	0.322	**0.162**
0.40	**0.202**	**0.202**	**0.202**	0.298	**0.202**	**0.216**	**0.216**	**0.216**	0.276	**0.216**
0.45	**0.227**	**0.227**	0.229	0.273	**0.227**	0.241	0.251	**0.237**	0.253	0.243
0.50	0.242	0.242	**0.236**	0.248	0.252	0.235	0.248	0.233	**0.230**	0.270
0.55	**0.223**	**0.223**	0.226	**0.223**	0.277	0.207	**0.200**	0.212	0.207	0.297
0.60	**0.198**	**0.198**	**0.198**	**0.198**	0.302	0.184	**0.182**	0.184	0.184	0.324
0.70	**0.149**	**0.149**	**0.149**	**0.149**	0.353	**0.138**	**0.138**	**0.138**	**0.138**	0.378

Notes: Table shows discrete loss values. The upper part relates to the SP500 index, the lower part – to the DM/USD exchange rate; the left half – to the in-sample computations, the right half – to the out-of-sample computations. The minimal criterion value(s) across each half of each line is in boldface.

Table 1c. In-sample and out-of-sample, weekly data

α	In-sample loss					Out-of-sample loss				
	Dynamic models			Trivial forecast		Dynamic models			Trivial forecast	
	D	R	S	1	0	D	R	S	1	0
					SP500w					
0.30	0.162	**0.148**	0.155	0.274	0.182	0.142	**0.128**	0.145	0.312	0.167
0.40	0.160	**0.151**	0.153	0.235	0.243	0.153	**0.131**	0.147	0.267	0.222
0.45	0.158	**0.148**	0.153	0.216	0.274	0.158	**0.138**	0.147	0.245	0.250
0.50	0.154	**0.142**	0.153	0.196	0.304	0.158	**0.123**	0.148	0.223	0.278
0.55	0.151	**0.136**	0.150	0.176	0.334	0.149	**0.126**	0.149	0.200	0.305
0.60	0.139	**0.135**	0.143	0.157	0.365	0.142	**0.129**	0.138	0.178	0.333
0.70	0.114	**0.110**	0.115	0.118	0.426	**0.113**	0.126	0.124	0.134	0.389
					DMUSDw					
0.30	**0.153**	0.156	0.154	0.342	0.154	0.174	**0.168**	**0.168**	0.308	**0.168**
0.40	**0.189**	**0.189**	0.201	0.293	0.205	0.211	0.207	**0.196**	0.264	0.224
0.45	0.199	**0.196**	0.202	0.268	0.230	0.219	0.218	**0.180**	0.242	0.252
0.50	**0.202**	0.207	**0.202**	0.244	0.256	**0.178**	0.185	**0.178**	0.220	0.280
0.55	0.207	0.203	**0.202**	0.220	0.282	0.188	0.190	**0.178**	0.198	0.308
0.60	0.187	**0.185**	0.195	0.195	0.307	0.177	0.184	**0.173**	0.176	0.336
0.70	0.148	**0.145**	0.146	0.146	0.358	0.135	**0.131**	0.132	0.132	0.392

Notes: Table shows discrete loss values. The upper part relates to the SP500 index, the lower part – to the DM/USD exchange rate; the left half – to the in-sample computations, the right half – to the out-of-sample computations. The minimal criterion value(s) across each half of each line is in boldface.

Table 2. Robustness check, daily SP500 data (SP500d)

α	Density models			Sign model	
	Constant s and k NGARCHSK	Varying s and k NGARCHSK	Normal GARCH(1,1)	Logit BARMA(1,1)	Logit BARMA(2,2)
			In-sample		
0.30	0.138	0.135	0.127	0.144	0.145
0.40	0.150	0.147	0.147	0.165	0.166
0.45	0.154	0.162	0.154	0.164	0.164
0.50	0.166	0.168	0.164	0.164	0.164
0.55	0.166	0.166	0.164	0.160	0.161
0.60	0.160	0.157	0.156	0.163	0.160
0.70	0.135	0.136	0.137	0.129	0.129
			Out-of-sample		
0.30	0.144	0.143	0.136	0.144	0.150
0.40	0.174	0.184	0.178	0.174	0.162
0.45	0.194	0.190	0.190	0.194	0.172
0.50	0.173	0.170	0.175	0.173	0.173
0.55	0.166	0.168	0.165	0.166	0.171
0.60	0.159	0.159	0.167	0.159	0.158
0.70	0.126	0.134	0.127	0.126	0.129

Notes: Table shows discrete loss values for various experiments with the SP500 index. The upper part relates to the in-sample computations, the lower part – to the out-of-sample computations.

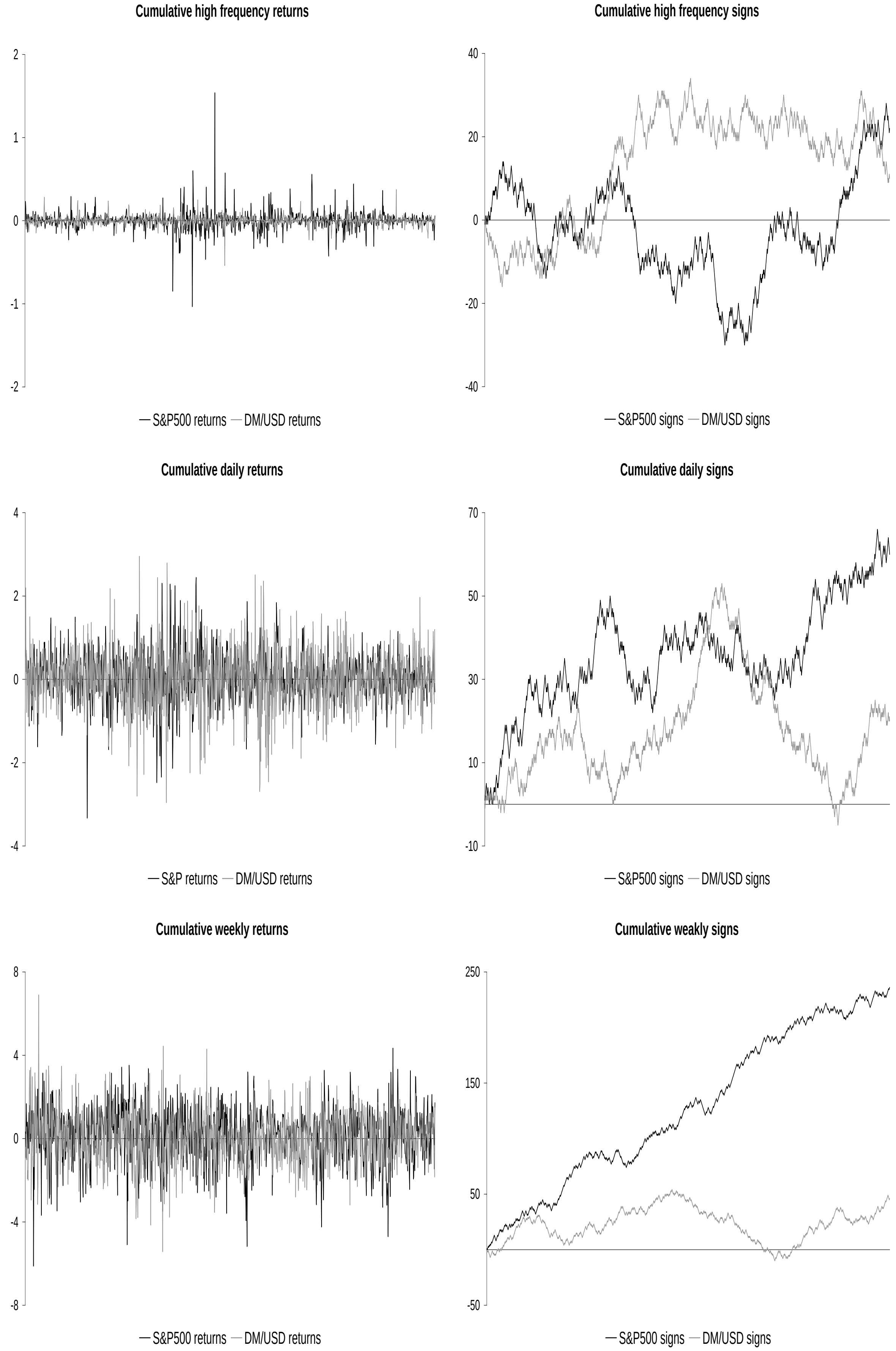

Figure 1: Data on returns and signs

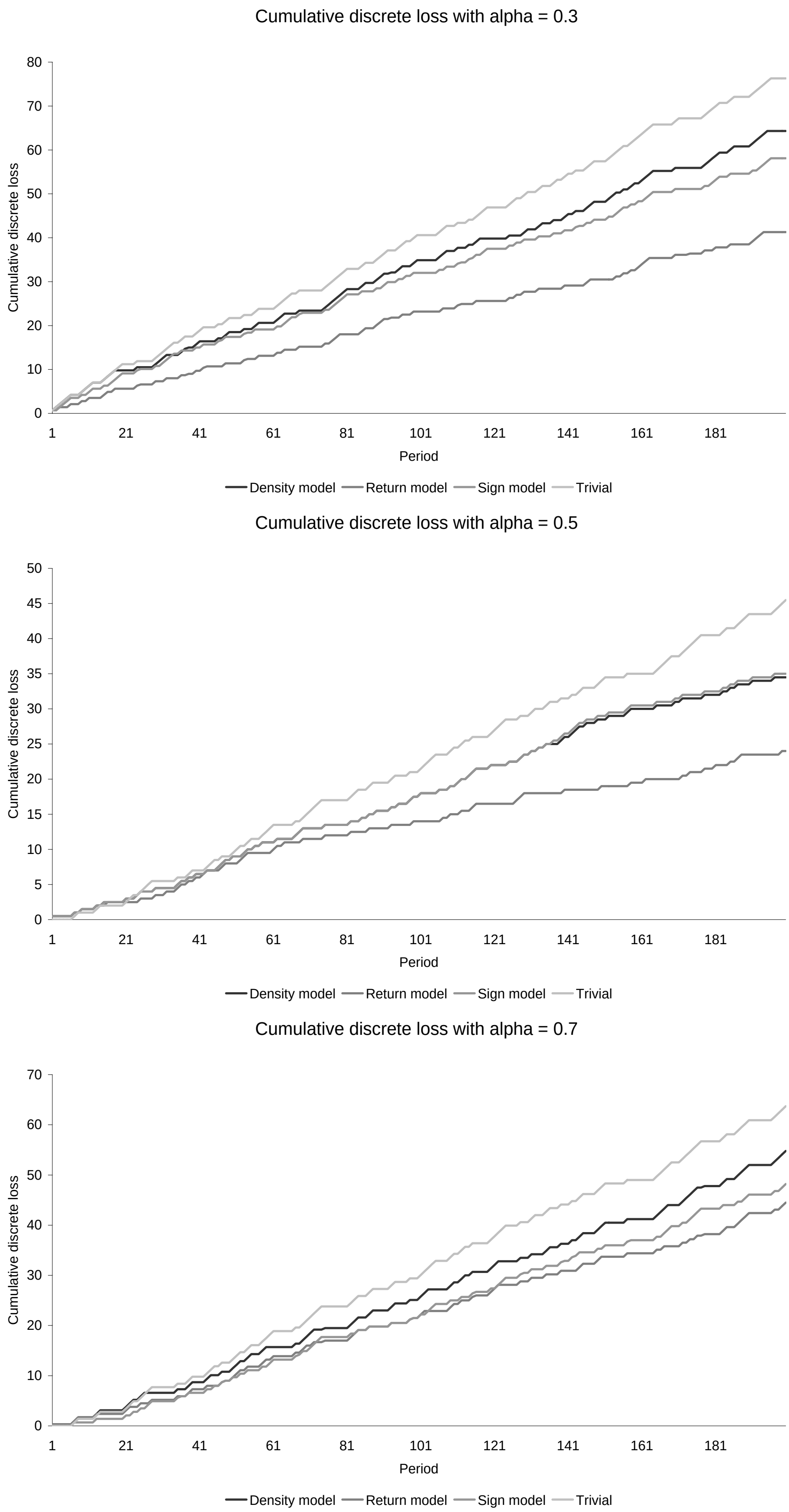

Figure 2: Out-of-sample forecasts for daily S&P500