Modeling the behavior of investors

Thomas Robin *  Michel Bierlaire *

February 23, 2011

Abstract

We propose an hybrid discrete choice framework for modeling decisions of investors performed on stock markets. We focus on the choice of action (buy or sell) and the duration until the next action. The choice of action is handled with a binary logit model with latent classes characterizing the perception of the risk, while a Weibull regression model is used for the duration until the next action. The duration model also accounts for the risk perception. Both models consider the dynamic nature of the underlying phenomenon. They are merged in a single model called combined model. It is estimated using data from a Swiss bank consisting of 25989 observations of transactions performed between January 2005 and September 2010, in 6 different funds. The predictive performance of the models are tested. A cross-validation analysis is performed. The forecasting accuracy of the action model is studied more in details. Parameters of both models are interpretable and emphasize interesting behavioral mechanisms related to investors’ decisions. The good predictive capabilities of the action model in a real context makes it operational.

1 Introduction

The prediction of the evolution of the stock market is crucial for investors in order to forecast their monetary gains. For authorities, this topic is important for regulating the market. The evolution of the stock market depends on the decisions taken by numerous financial actors, including asset managers, firms, long-term and short-term investors, or unprofessional individuals. In addition, automatic trading based on algorithms is

*Transport and Mobility Laboratory, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland, {thomas.robin, michel.bierlaire}@epfl.ch
also used by banks for taking advantage of the instantaneous price variations of stocks. A recent study done by the Aite Group (Aite, 2010) shows that in 2009, slightly more than half of European equity volume is executed electronically. They forecast an increase of this share in a near future. The actors’ decisions are based on information coming from very eclectic sources and vary across actors. Information about the stock such as the price, play a key role, as well as information associated with the underlying company and the stock market. News and governmental announcements are considered in these decisions. As the process is dynamic, the information of the previous days are relevant. The variations of the information reflect the interaction between the different actors’ decisions, according to the supply and demand rule. But the stock market is a complex system, hard to understand and model.


Since the early 90’s the behavioral finance has a growing interest. The evolution of financial market is caused by the behavior of the underlying actors, which appears to be

Different methodologies have been applied and developed. Specific time series techniques are common. Bollerslev et al. (1992) review theory and applications of ARCH model in finance. Mikosch and Štěrčák (2004) develop models in order to account for non stationarity in financial time series. Rule-based models are largely used. Zopounidis (1999) study the contribution of multicriteria analysis in solving financial decision problems in a realistic context. Wagner et al. (2002) propose a guidance for specifying rules of expert systems in the financial field. Operation research techniques appear to be useful for optimization. El-Yaniv (1998) survey results concerning online algorithms for solving problems related to the management of money. Machine learning methods are well adapted to the financial context, due to the huge amount of aggregate data available on financial markets. Giles et al. (2001) develop neural networks (NN) for predicting noisy financial time series. jae Kim (2003) use SVN to forecast financial time series. Huang et al. (2004) compare artificial intelligence techniques for the credit rating analysis (SVN against NN). Other methods have been used. For example, Fer- manian and Scaillet (2004) discuss the use of copulas for modeling cross-dependences in financial applications, or Embrechts and Schmidli (1994) review methodologies used to model stochasticity in finance.

In this paper, we investigate the development of behavioral models designed to capture the behavior of specific actors in the financial sector. In that purpose, we propose to use an hybrid discrete choice framework to model the behavior of investors. Discrete choice models (DCM) are well adapted in the context of disaggregate modeling (Ben-Akiva and Lerman, 1985). They have the advantage to explicitly capture causal effects. Interestingly, few articles report the use of DCM in the financial context. Johnsen and Melicher (1994) and Hensher and Jones (2004) developed DCM for predicting the firm financial distress. de Palma et al. (2008) consider the inclusion of the risk perception
in random utility models, although they do not focus on finance per se.

We are interested in modeling the behavior of investors working in banks with long-term equity investments. Each fund is managed by a person, who is in charge of a team. The decisions are not always taken by the fund manager himself as individuals of the team are also taking decisions. The fund manager validates these decisions. The fund gathers companies per thematic. It could be the size, the activity sector, the nationality, or a mix between these thematics. Each day and regarding stocks, an investor takes financial decisions. His purpose is to adjust his portfolio in order to maximize the returns. For each stock within the fund, he decides to buy, sell or wait. In case of buying and selling, he has to decide the involved quantity of money. Then, the order is passed to the trading service who is in charge to implement the decision. The trader translates the decision into transactions. Several transactions can reflect the same decision as the trader chooses the right moments in the day for taking advantage of the market, or for reducing the impact of the decision on the associated stock.

The investor considers several and heterogeneous information sources. The stock market gives information about the stocks, and official data are reported by the companies. In addition, the bank provides indicators for the stocks and the underlying companies. They reflect their current states and evolutions and are based on public information. The precise formula of these indicators remains secret. Indices represent the market state and the market risk. Other information such as news about politics, fusion of companies or governmental announcements are relevant. The investor considers the money flows within the bank. If bank clients feed their account, the money enters and the investor is obliged to buy stocks to make the money grow. If bank clients close their account, the money leaves and the investor has to sell stocks for generating cash.

We focus on two behavioral aspects which are the choice of action (buy or sell) and the duration between two actions. Regarding the choice of action, the developed model is inspired from the work of Walker (2001) and Greene and Hensher (2003) about DCM with latent classes. In our case, the latent classes account for the risk perception. For the duration, we considered the models presented by den Berg (2001) and Bauwens and Veredas (2004). A Weibull model appears to be appropriate. It accounts for a different risk perception than the model of action choice.

This paper is organized into various sections as follows. In Section 2, we present the raw data, in Section 3 the notations, in Section 4 the time discretization, in Section 5 the explanatory variables. Section 6 details the model. In Section 7, the estimation results are shown. Section 8 validates the model.
2 Raw data

We have access to transactions initiated by internal investor of the Swiss bank Lombard Odier Darier Hentsch, in six funds. For confidentiality reasons, the data are anonymous. Funds 2 and 3 are managed by the same person, as well as funds 4 and 5.

The raw data consists in 25989 observations of transactions passed by traders. The time period goes from 2005.01.03 to 2005.09.13. Each transaction is characterized by a date, a company, a fund and an amount of money. The amount is positive if stocks have been bought and negative if they have been sold. Stocks of 1236 companies are considered. The number of observations and companies per fund are shown in Table 1. Some companies appear in several funds. The shares between the transactions buy and sell are also shown. The transactions are equally spread between buy and sell for all the funds, except fund 3.

<table>
<thead>
<tr>
<th>Fund</th>
<th>Nb of transactions</th>
<th>Nb of companies</th>
<th>% buy</th>
<th>% sell</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4354</td>
<td>160</td>
<td>44.63</td>
<td>55.37</td>
</tr>
<tr>
<td>2</td>
<td>1189</td>
<td>64</td>
<td>55.82</td>
<td>47.18</td>
</tr>
<tr>
<td>3</td>
<td>6427</td>
<td>363</td>
<td>78.70</td>
<td>21.30</td>
</tr>
<tr>
<td>4</td>
<td>2018</td>
<td>560</td>
<td>54.21</td>
<td>45.79</td>
</tr>
<tr>
<td>5</td>
<td>6935</td>
<td>55</td>
<td>45.84</td>
<td>54.16</td>
</tr>
<tr>
<td>6</td>
<td>5066</td>
<td>185</td>
<td>57.26</td>
<td>42.74</td>
</tr>
</tbody>
</table>

Table 1: Number of transactions, number of companies and percentage of actions per fund in the raw data

The data are dynamic per nature. The evolution of the number of transactions per fund is presented in Figure 1. The number of buy transactions is stable in average, but with strong fluctuations. Regarding sell transactions, two phases appear. The first goes from 2005 to 2008, which is characterized by stability and weak fluctuations. The second goes from 2008 to 2010, it is stable in average but with stronger fluctuations compared to phase 1. The average in phase 2 is higher than in phase 1. These two observations are linked to the fact that the starting of phase 2 coincides with the financial crisis of 2008.

Five fundamental indicators are considered as explanatory variables. They are derived by financial analysts of the bank from several public indicators. They are called quality, sentiment, technic, value and price. Quality measures the fundamental quality of a company by examining specific economic and financial data published by the company. Sentiment is a number based on a combination of estimates of the analysts
Figure 1: Evolution of the number of actions *buy* and *sell* contained in the raw data covering the company, as for example the next year earnings estimates. *Technic* is a combination of indicators that analyze the company's activity on the market by identifying chart pattern of prices, as for example the momentum reversal. The momentum is defined by the difference between two prices of a same stock for a chosen time horizon. *Value* is an objective value of a company based on current valuation metrics, like for
example price to earning ratio. Price characterizes the price of the stock associated to the company. Note that portfolio information is not available in the data, neither data about investor himself.

For a certain company, we display the variations of the five indicators in Figure 2. These values are scores and have no unit. For the sake of reading, the values have been normalized between 0 and 1 by adding the observed minimum per fund and dividing by the observed maximum per fund. A correlation analysis performed between the different indicators did not evidence strong links between them. This is logical because, they have been built to reflect complementary information about the company. The quality and sentiment have a tendency to be constant over small time period, that's why we observe levels on their associated curves (see Figures 2(a) and 2(e)). This is not the case for the other indicators which present continuous variations. All these indicators are positively correlated with the financial health of the company.

The decision context depends on the state of the stock market. It is characterized by market indices, easily accessible. The VIX (symbol for the Chicago board options exchange market volatility index) is retained. This is a popular measure of the implied volatility of the S&P 500 index, which reflects the average price level of the stock market in United States. Consequently the VIX characterizes the market risk. It has been plotted for the considered period of time in Figure 3. The horizontal line is set for the VIX equal to 25. According to investors, this value shows the limit between two complementary decisional behaviors. The financial crisis of the end 2008 appears clearly. It is characterized by the highest pick on the curve.

3 Notations

We introduce the notations that are used along the paper, specially in Sections 4, 5 and 6. They are used and not redefined afterward. The modeling concepts are defined specifically in Section 6.

- DCM: discrete choice models;
- i.i.d.: independent and identically distributed;
- EV(0, 1): standardized extreme value distribution;
- N(0, 1): standardized normal distribution;
- c: company;
- C: the number of companies;
Figure 2: Examples of fundamental indicators for one company

- $t$: day;
- $T$: the length of the entire time period;
- $t_{14}$: time horizon in days;
- $H$: vector of considered $t_{14}$, $H = \{1, 5, 10, 15, 20, 25, 30, 60, 90, 180, 270, 360\}$;
- $f$: fund, $f \in F = \{1, 2, 3, 4, 5, 6\}$;
- $F$: vector of funds;
- $g$: group of funds, $g = 1$ groups funds 1, 2, 3, $g = 2$ groups funds 4, 5, 6;
- $t' = t + D(c, t) + 5$: the day of the action performed after $A(c, t)$, $A(c, t)$ and $D(c, t)$ are defined in Section 3.1;
Figure 3: Evolution of the VIX during the considered period of time

- B: action buy;
- S: action sell;
- R: risky situation;
- N: normal situation, as opposed to R;
- c: experience of the cross-validation;
- t_{0,c}: starting date of the simulation set of experience c;
- T_c: ending date of the simulation set of experience c;
- R^2_c: R^2 predicted by the duration model on the simulation set of experience c;
- z_{c,t}: standardized residual of the duration model.

Remaining notations are organized per thematics.
3.1 Variables

Notations for dependent and explanatory variables are summarized below.

- \(O_{c,t}\): transaction observed on \(t\) for \(c\) (money), if it is positive stocks have been bought, if it is negative stocks have been sold;
- \(A(c,t)\): action decided on \(t\) for \(c\), \(A(c,t) \in \{B, S\}\);
- \(D(c,t)\): time duration between \(A(c,t)\) and \(A(c,t')\) in weeks (5 days);
- \(\hat{D}(c,t)\): predicted duration;
- \(\bar{D}(c,t)\): duration mean calculated over \(t\) and \(c\).
- \(r_A\): risk in the action model, \(r_A \in \{N, R\}\);
- \(r_D\): risk in the duration model, \(r_D \in \{N, R\}\);
- \(\text{qual}_{c,t}\): quality associated to \(c\) on \(t\);
- \(\text{tech}_{c,t}\): technic associated to \(c\) on \(t\);
- \(\text{sent}_{c,t}\): sentiment associated to \(c\) on \(t\);
- \(\text{pric}_{c,t}\): price associated to \(c\) on \(t\);
- \(\text{valu}_{c,t}\): value associated to \(c\) on \(t\);
- \(x_{c,t} = \{\text{qual}_{c,t}, \text{tech}_{c,t}, \text{sent}_{c,t}, \text{pric}_{c,t}, \text{valu}_{c,t}\}\): vector containing the 5 fundamental indicators for \(c\) on \(t\);
- \(K_{c,t}\): the length of \(x_{c,t}\)
- \(VIX_t\): VIX on \(t\);
- \(\text{Perf}(x_{c,t}(k), t_H)\): performance of \(x_{c,t}(k)\), calculated on \(t_H\) (Equation (1));
- \(\text{Long}(x_{c,t}(k), t_H)\): long-term value of \(x_{c,t}(k)\), calculated on \(t_H\) (Equation (2));
- \(\text{Short}(x_{c,t}(k), t_H)\): short-term value of \(x_{c,t}(k)\), calculated on \(t_H\) (Equation (3));
- \(\text{Sigm}(x_{c,t}(k), t_H)\): standard-error of \(x_{c,t}(k)\), calculated on \(t_H\) (Equation (4));
- \(X_{c,t}\): vector of raw and transformed values of \(\{x_{c,t}(k)\}_{k=1...K_{c,t}}\) (Equation (5));
- \(Y_t\): vector of raw and transformed values of the \(VIX_t\) (Equation (6)).

3.2 Parameters

Two models are introduced in Section 6. The parameters of the models are denoted as follows.
3.2.1 The action model

- $\beta$: vector of parameters (Equation (24));
- $\mu$: scale of the random variables $\varepsilon_{B_rA,c,t'}, \varepsilon_{S_rA,c,t'}$;
- $\omega_{A_r}$: vector of parameters associated to the risk perception (Equation (8));
- $\beta_{B_r}$: vector of parameters associated to the explanatory variables;
- $K_B$: size of $\beta_B$;
- $\text{ASC}_{B,rA}$: constant parameter;
- $\alpha_{B,rA}$: parameter associated to the deterministic utility of B in the previous action;
- $\lambda_{B,rA}$: parameter weighting the influence of the deterministic utility of B in the previous action.

3.2.2 The duration model

- $\theta$: vector of parameters (Equation (28));
- $\eta_D$: shape parameter of the Weibull distribution;
- $\theta_D$: vector of parameters associated to the explanatory variables;
- $K_D$: size of $\theta_D$;
- $\omega_D$: vector of parameters associated to the risk perception (Equation (13));
- $\text{ASC}_{D,rD,g}$: constant parameter;
- $\alpha_{D,rD,g}$: parameter capturing the effect of the previous duration;
- $\theta_{B,rD}$: parameter capturing the influence of the deterministic utility of B.

4 Time discretization

In the raw data, we do not observe the investors’ decisions but direct consequences of them. For a given stock, once the investor has decided an action and an associated amount of money, a trader implements the decision. Traders have a tendency to split the investors’ decisions in several successive transactions in order to decrease the influence of the decisions on the underlying stocks, in terms of price, due to the supply and demand rule. These transactions constitute the data. As a consequence, different transactions in the raw data reflect the same decision. The date of the decision and the date of the first transaction coincides. Transactions have to be aggregated for each
stock in order to represent the investors’ decisions.

For given stock, we group transactions in sets. A transaction belongs to a set if at least one transaction inside the set is separated from the considered transaction from less than five days. Then, the time period of each set is split in consecutive time windows of five days. The transactions within the time windows of five days constitute subsets. Within subsets, the transactions are aggregated by summing their associated amounts of money. If the sum is positive, it is a buy action; if it is negative, a sell action appears. The information of the first day of the subset are the explanatory variables. Actions involving a small quantity of money have been discarded.

A time window of 5 days is used, because it corresponds to a working week. This has been validated by the involved investors. Two situations appear when aggregating the transactions, they are illustrating in the following:

1. A transaction with no near neighbor. This means that the set contains only one transaction. It is considered as an investor’s decision. An example of this situation is presented in Figure 4. In that case, there is no aggregation.

\[ O_{c,t} \quad O_{c,t+8} \]

\[ \uparrow \quad \uparrow \]

\[ A(c, t) \quad A(c, t + 8) \]

days

Figure 4: Aggregation of isolated transactions

2. A time period with neighboring transactions. This means that a set contains at least two transactions. The time period is covered by non-overlapping time windows of five days. Transactions are aggregated within each subset and information about the first date of the subset are considered. An example is shown in Figure 5.

Note that a minimum of five days separate two actions. Actions with small quantities of money have been removed because they correspond to adjustments representing noise in this modeling context. According to investors, they concern 25\% of the actions. After processing the transactions, the data contains 9178 observations of actions performed on stocks of 1121 companies. The details are presented in Table 2. The shares between actions are equally distributed and rather the same than for the transactions presented in Table 1, except for the fund 3. For this fund, the distribution is much...
Figure 5: Aggregation of neighboring observations

more balanced between the actions, compared to the transactions.

<table>
<thead>
<tr>
<th>Fund</th>
<th>Nb of decisions</th>
<th>Nb of companies</th>
<th>% buy</th>
<th>% sell</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1461</td>
<td>145</td>
<td>54.96</td>
<td>45.04</td>
</tr>
<tr>
<td>2</td>
<td>913</td>
<td>58</td>
<td>53.34</td>
<td>46.66</td>
</tr>
<tr>
<td>3</td>
<td>508</td>
<td>50</td>
<td>59.45</td>
<td>40.55</td>
</tr>
<tr>
<td>4</td>
<td>3738</td>
<td>505</td>
<td>51.66</td>
<td>48.34</td>
</tr>
<tr>
<td>5</td>
<td>1659</td>
<td>316</td>
<td>43.40</td>
<td>56.60</td>
</tr>
<tr>
<td>6</td>
<td>899</td>
<td>175</td>
<td>45.05</td>
<td>54.95</td>
</tr>
</tbody>
</table>

Table 2: Number of actions, number of companies and percentages of actions per fund in the processed data

The repartition of the actions across time is presented in Figure 6. Compared to Figure 1, the aggregation has been quite strong for the buy transactions from 2005 to 2007. Lots of situations 2 appeared (see Figure 5). Regarding other periods, the graphs are similar, showing the higher propensity of situations 1 (see Figure 4).

5 Explanatory variables

Investors account for the dynamic of the variables (Figures 2 and 3) when taking decisions. New variables have been computed based on the five fundamental indicators and the VIX for reflecting this dynamic. The framework of the dynamic data calculation is presented in Figure 7. Two consecutive actions are represented \( A(c, t) \) and \( A(c, t + 5) \). As explained in Section 4, two consecutive actions are separated by a minimum of five days.

The dynamic variables are calculated as follows. We call performance the relative variation
Figure 6: Evolution of the number of actions *buy* and *sell* contained in the processed data

\[
\text{Perf}(x_{c,t}(k), t_H) = \frac{x_{c,t}(k) - x_{c,t-t_H}(k)}{x_{c,t-t_H}(k)}.
\]

The mean calculated over \( t_H \) is called the *long-term value*.
\text{Long}(x_{c,t}(k), t_H) = \frac{1}{t_H} \sum_{t=t_h}^{t} x_{c,t}(k). \hspace{1cm} (2)

The difference between the current value \(x_{c,t}(k)\) and the long-term value \(\text{Long}(x_{c,t}(k), t_H)\) is called the \textit{short-term value}

\text{Short}(x_{c,t}(k), t_H) = x_{c,t}(k) - \text{Long}(x_{c,t}(k), t_H). \hspace{1cm} (3)

Finally, we define the \textit{standard-error} as

\text{Sigm}(x_{c,t}(k), t_H) = \sqrt{\frac{1}{t_H} \sum_{t=t_h}^{t} (x_{c,t}(k) - \text{Long}(x_{c,t}(k), t_H))^2}. \hspace{1cm} (4)

It characterizes the variations of \(x_{c,t}(k)\) within \(t_H\). We explicit

\[ X_{c,t} = \{ x_{c,t}(k), \text{Perf}(x_{c,t}(k), t_H), \text{Long}(x_{c,t}(k), t_H), \text{Short}(x_{c,t}(k), t_H), \text{Sigm}(x_{c,t}(k), t_H) \}_t \in H, k = 1 \ldots k, \]  

and

\[ Y_t = \{ \text{VIX}_t, \text{Perf}(\text{VIX}_t, t_H), \text{Long}(\text{VIX}_t, t_H), \text{Short}(\text{VIX}_t, t_H), \text{Sigm}(\text{VIX}_t, t_H) \}_t \in H, \]  

Heterogeneous \(t_H\) are considered, \(t_H \in H = \{ 1 \ldots 5, 10, 15, 20, 25, 30, 60, 90, 180, 270, 360 \} \). This is motivated by the fact that investors consider short-term to long-term dynamic of variables when making decisions. 366 variables are considered in total (5 indicators and the VIX, 4 transformations, 15 time horizons, \(366 = 6 \times (1 + 4 \times 15) \)).
The variables have been normalized per fund (except the VIX) by sequentially subtracting the minimum and dividing by the maximum. Then the variables are in $[0,1]$. The normalization has been done per fund because investors are considering the entire fund when taking decisions. The VIX has not been normalized because it is not fund specific, its interval of variation is manageable, and for interpretation of the proposed models (see Section 6).

For the considered time period, examples of the evolution of the dynamic variables are presented in Figure 5. The raw indicator is the price shown in Figure 8(a), which is the same than in Figure 2(d). $t_{H} = 1$ for the calculation of the performance, long-term value and short-term value. The performance and short-term value capture the immediate variations of the variable (Figures 8(b) and 8(d)), whereas the long-term value is the smoothed version of the raw variable (Figure 8(c)). Note that the smoothing degree is $t_{H}$. Due to the small value of $t_{H}$, it is qualitatively similar to the variation of the raw variable (Figure 8(a) and Figure 8(c)). The evolution of the standard-error is the most regular, because $t_{H} = 60$ for its calculation.

A correlation analysis has been performed between the action variable $A(c,t)$ and the explanatory variables. Results are summarized in Table 3. B is coded 0, and S is coded 1. If the correlation is positive sell is favored, otherwise it is buy. Note that a Pearson test has been performed for each correlation. Only significant effects have been kept. The correlations are not very high, the highest value is 0.449 and is observed for the standard-error of the VIX for $t_{H} = 360$, in fund 3. The number of significant and generic correlations across funds is low, compared to the number of variables (366). This points out the difference of financial management between funds, and emphasizes the specificity of the investors’ behavior within each fund. Nevertheless, generic and significant correlations provide information. Regarding variables associated with companies and stocks, time horizons are low, showing the propensity of investors to account for immediate information in their decisions. A difference of behavior appears between funds 1, 2, 3 and 4, 5, 6. Except for the variables associated to the VIX, correlations have the same signs within the two fund groups, and are opposed between the two groups. This difference is partly explained by the fact that in the two funds groups, an investor manage two funds, as explained in the Section 2 (funds 2 and 3 are managed by the same investor, as well as funds 4 and 6).

A correlation analysis has been also performed by splitting the processed data into two parts according to the level of VIX. This has been done in order to test if there is a significant difference of behavior in volatile and non-volatile markets. The considered threshold is 25 (see Figure 3). In case of low VIX, the significant and generic correlations are the same than in Table 3. Interpretations remain the same. In case of high VIX, only the variable Short($valu_{c,t},60$) stands out. This underlines the specificity of the investors’ behavior within each fund in risky situations. This is logical, because in
these situations, individualities predominate rules. Correlations are displayed in Table 4. The difference between the two groups of fund appear and is consistent with Table 3. The time horizon is higher (60 days compared to 3 days), which is logical. In risky situations, investors have more tendency to consider long-term information.

These correlation analysis help us to get intuition about the data, but are limited due to their univariate nature.

6 Model specification

We aim at understanding and modeling the financial decisions of an investor in a given time horizon (expressed in days), regarding a set of stocks. Given that an action is
Table 3: Generic and significant correlations between the action variable and the explanatory variables

<table>
<thead>
<tr>
<th>Transform</th>
<th>Variable</th>
<th>$t_H$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perf()</td>
<td>Price</td>
<td>1</td>
<td>0.065</td>
<td>0.139</td>
<td>0.127</td>
<td>-0.279</td>
<td>-0.238</td>
<td>-0.329</td>
</tr>
<tr>
<td>Short()</td>
<td>Value</td>
<td>3</td>
<td>-0.090</td>
<td>-0.084</td>
<td>-0.099</td>
<td>0.227</td>
<td>0.170</td>
<td>0.277</td>
</tr>
<tr>
<td>Sigm()</td>
<td>VIX</td>
<td>360</td>
<td>0.126</td>
<td>0.303</td>
<td>0.449</td>
<td>-0.062</td>
<td>-0.114</td>
<td>0.213</td>
</tr>
<tr>
<td>Short()</td>
<td>Technic</td>
<td>1</td>
<td>-0.065</td>
<td>-0.110</td>
<td>-0.113</td>
<td>0.213</td>
<td>0.181</td>
<td>0.257</td>
</tr>
</tbody>
</table>

Table 4: Generic and significant correlations between the action variable and the explanatory variables, for a high level of VIX

performed on day $t$ for stocks of company $c$, we assume that the investor decides the type of action (buy or sell), and the duration until the next action performed on the same stock. The decision about the duration is not supposed to be revised, once it has been taken. An overview of the decision process is shown in Figure 9. $D(c, t)$ is the duration between $A(c, t)$ and the next action $A(c, t')$, with $t' = t + D(c, t) + 5$. If $D(c, t) = 0$, the duration between the two consecutive actions is five days, which is the minimum duration according to the aggregation presented in Section 4. We model the decisions taken in $t'$, $A(c, t')$ and $D(c, t')$, conditionally on $A(c, t)$ and $D(c, t)$.

![Figure 9: The process of investors' decisions](image)

18
The model associated to \( A(c, t') \) is called action model. The model associated to \( D(c, t') \) is called duration model. The combined model groups the action and duration models. The general modeling framework is presented in Figure 10. Square shapes represent observed variables, whereas round shapes are latent variables. Shapes with dotted lines are for random variables, whereas plain lines are associated to deterministic variables. Arrows stand for causal links between variables, each arrow is associated to an equation. Plain arrows stand between variables of \( t' \), dotted arrows link variables in \( t \) to variables in \( t' \). We define the modeling concepts of the scheme per models. For the action model, we specify:

- \( V_A(R, c, t'|\omega_D) \): the measure for the risk \( R \) in the action model (Equation (7));
- \( W_A(r, c, t'|\omega_A) \): the randomized measure of the market risk \( r_A \) (Equation (9));
- \( V_B(c, t'|r_A, \beta) \): the deterministic utility associated with the alternative B (Equation (18));
- \( M_B(c, t'|r_A, \beta) \): the term capturing the effect of the previous action on the current choice of action (Equation (19));
- \( U_B(c, t'|r_A, \beta) \): the random utility of the alternative B (Equation (17));
- \( U_S(c, t'|r_A, \beta) \): the utility of the alternative S (Equation (17));
- \( \varepsilon_{A,r_A,c,t'} \): a random variable associated to the risk \( r_A \) (Equation (10));
- \( \varepsilon_{B,r_A,c,f,t'} \): a random variable associated to B, under \( r_A \) (Equation (20));
- \( \varepsilon_{S,r_A,c,f,t'} \): a random variable associated to S under \( r_A \) (Equation (20)).

For the duration model we specify:

- \( V_D(R, t'|\omega_D) \): the measure for the risk \( R \) (Equation (12));
- \( W_D(r_D, t'|\omega_D) \): the randomized measure of the market risk \( r_D \) (Equation (14));
- \( m_D(c, t'|r_D, \theta, \beta) \): a utility (Equation (27));
- \( \lambda_D(c, t'|\theta, \omega_D, \beta) \): scale parameter of the Weibull distribution (Equation (29));
- \( \varepsilon_{D,r_D,t'} \): a random variable associated to the risk \( r_D \) in the duration model (Equation (15));
- \( \varepsilon_{D,t} \): a random variable, \( D(c, t) \) is assumed to be its mean (Equation (25)).

Descriptive statistics shown in Table 3 allow to underline a significant difference of behavior between investors managing funds 1, 2, 3 and 4, 5, 6. We account for this difference in the specification of the models. In Section 6.1 the risk perception in both models are detailed. In Section 6.2 the action model is presented, and in Section 6.3 it is the duration model.
6.1 The risk perception

Day and Huang (1990) define three market types: bear, bull and sheep markets. A bear market corresponds to a decreasing confidence of the investors in the market, generating an increase of the risk in terms of returns. This is the contrary for the bull market. The sheep market represents an intermediary position between the bull and bear markets. In that case, the majority of the investors follow the market tendencies. According to the investors implicated in the observed decisions, two types of behavior occur depending on the market risk. The first behavior is called normal, corresponding to the bull and sheep markets. The second is called risky corresponding to the bear market. The risk perception is not directly observed, and has a strong influence on the investors’ decisions. Two models for the risk perception have been developed, one associated to the action model (see Section 6.1.1) and one to the duration model (see Section 6.1.2). No characteristics of the investors were available in the data, so the risk perception only depends on attributes of the decision context.

6.1.1 The risk perception in the action model

The risk \( r_A \) is a discrete variable. A model for risk classification is developed. A logit function is used. The deterministic measure of the risk \( R \) is

\[
V_A(R, c, t' | \omega_A) = ASC_{W_A} + \omega_{A,1}VIX_tI_{c,g=1} + \omega_{A,2}VIX_tI_{c,g=2} + \omega_{A,3}\text{Sigm}(\text{Sent}_{c,t'}, 5),
\]

where \( I_{c,g} \) is an indicator equal to 1 if \( c \) belongs to \( g \), 0 otherwise. We explicit

\[
\omega_A = \{ASC_{W_A}, \omega_{A,1}, \omega_{A,2}, \omega_{A,3}\}.
\]

The randomized measure of risks \( N \) and \( R \) are

\[
W_A(N, c, t' | \omega_A) = \varepsilon_{A,N,c,t'},
\]

\[
W_A(R, c, t' | \omega_A) = V_A(R, c, t' | \omega_A) + \varepsilon_{A,R,c,t'},
\]

and assuming

\[
\varepsilon_{A,r_A,c,t',i,j} \sim \text{EV}(0, 1), \text{ for } r_A \in \{N, R\},
\]

a binary logit model is derived where the alternatives are the risks \( N \) and \( R \). The associated probabilities are

\[
P_A(N, c, t' | \omega_A) = \frac{1}{1 + e^{V_A(R, c, t' | \omega_A)}},
\]

\[
P_A(R, c, t' | \omega_A) = \frac{1}{1 + e^{-V_A(R, c, t' | \omega_A)}},
\]

20
We expect \( V_A(R, c, t'|\omega_A) \) (Equation (7)) to increase when the risk increases, so \( \omega_{A,1}, \omega_{A,2} \) and \( \omega_{A,3} \) should be positively estimated. The risk perception depends both on the market and on \( c \), due to presence of the VIX and \( \text{Sigm}(\text{Sent}_{c,t'}, 5) \) (attribute described in equation 4).

### 6.1.2 The risk perception in the duration model

The risk \( r_D \) is a discrete variable. A model for risk classification is developed. A logit function is used. The deterministic measure of the risk \( R \) is

\[
V_D(R, t'|\omega_D) = ASC_{W_D,1}I_{c,g=1} + ASC_{W_D,2}I_{c,g=2} + \omega_{D,1}\text{VIX}_{t'}I_{c,g=1} + \omega_{D,2}\text{VIX}_{t'}I_{c,g=2},
\]

where \( I_{c,g} \) is an indicator equal to 1 if \( c \) belongs to \( g \), 0 otherwise. We explicit

\[
\omega_D = \{ASC_{W_D}, \omega_{D,1}, \omega_{D,2}\}
\]

The randomized measure of the risks \( N \) and \( R \) are

\[
W_D(N, t'|\omega_D) = \varepsilon_{D,N,t'},
W_D(R, t'|\omega_D) = V_D(R, t'|\omega_D) + \varepsilon_{D,R,t'},
\]

and assuming

\[
\varepsilon_{D,r_D,t'} \sim \text{EV}(0, 1), \text{ for } r_D \in \{N, R\},
\]

a binary logit model is derived where the alternatives are the risks \( N \) and \( R \). The associated probabilities are

\[
P_D(N, t'|\omega_D) = \frac{1}{1 + e^{V_D(N, t'|\omega_D)}},
\]

\[
P_D(R, t'|\omega_D) = \frac{1}{1 + e^{-V_D(R, t'|\omega_D)}},
\]

We expect \( V_D(R, t'|\omega_D) \) (Equation (12)) to increase when the risk increases, so \( \omega_{D,1}, \omega_{D,2} \) should be positively estimated. Contrary to \( r_A \) (Section 6.1.1), The risk perception depends only on the market.

### 6.2 The action model

The choice of action is a discrete choice situation. We develop a binary logit model with two latent classes corresponding to the two risk situations. The random utilities of the two alternatives \( B \) and \( S \) are
\[ U_B(c, t'|r_A, \beta) = V_B(c, t'|r_A, \beta) + M_B(c, t'|r_A, \beta) + \epsilon_{B,r_A,c,f,t'}, \]
\[ U_S(c, t'|r_A, \beta) = \epsilon_{S,r_A,c,f,t'}, \] (17)

with

\[ V_B(c, t'|r_A, \beta) = \lambda S C_{B,r_A} + \sum_{g \in \{1,2\}} I_{g,c} \beta_{B,k} \sum_{l=1}^{K_X} I_{B,g,k,l,r_A} X_{c,t'}(l) \] (18)

where \( I_{B,g,k,l,r_A} \) is an indicator equal to 1, if the parameter \( \beta_{B,k} \) is associated to the attribute \( X_{c,t'}(l) \), to the group of fund \( g \) and appear under the risk \( r_A \). \( I_{g,c} \) is an indicator equal to 1 if \( c \) belongs to \( g \).

\[ M_B(c, t'|r_A, \beta) = \alpha_{B,r_A} V_B(c, t|r_A, \beta)e^{\lambda A,r_A D(c,t)} \] (19)

is a term accounting for the effect of the previous action, represented by the deterministic utility of \( B \) in the previous action, weighted by \( D(c, t) \), the duration between the last and the current action performed on stocks of \( c \). The assumptions about the random terms are

\( \epsilon_{B,r_A,c,f,t'}, \epsilon_{S,r_A,c,f,t'} \) i.i.d. \( EV(\mu, \mu) \), for \( r_A \in \{N, R\} \), (20)

The probabilities of the actions \( B \) and \( S \) under the risk \( r_A \) are

\[ P_B(c, t'|r_A, \beta) = \frac{1}{1 + e^{-\mu V_B'}}, \]
\[ P_S(c, t'|r_A, \beta) = 1 - P_B(c, t'|r_A, \beta), \] (21)

where

\[ V_B' = V_B(c, t'|r_A, \beta) + M_B(c, t'|r_A, \beta). \] (22)

After having summed on the risks \( N \) and \( R \), the probabilities of the actions come

\[ P_B(c, t'|\beta, \omega_A) = P_B(c, t'|N, \beta)P_A(N, c, t'|\omega_A) + P_B(c, t'|R, \beta)P_A(R, c, t'|\omega_A), \]
\[ P_S(c, t'|\beta, \omega_A) = 1 - P_B(c, t'|\beta, \omega_A), \] (23)

where \( P_A(N, c, t'|\omega_A), P_A(R, c, t'|\omega_A) \) are the probabilities to be in the risk \( R \) defined in equation (11). The vector of parameters \( \beta \) is then
\[ \beta = \{ (\beta_{B,k})_{k=1}^{K,0}, \{ \alpha_{B,r}, \lambda_{B,r} \}_{r=N,R}, \{ \mu_l \}_{l=2,...,6} \}. \] (24)

Only attributes for \( t_H = 1 \) are used in the deterministic utility shown in Equation (18). This is explained by the statistical analysis presented in Table 3. Dynamic variables calculated with small \( t_H \) are significantly correlated with the action choice. In a multivariate context \( t_H = 1 \) appear to be the most appropriate (see Section 7). We expect \( \lambda_{B,r} \) to be negative, as we suppose the impact of the previous action to decrease when the duration between the previous and the current action increases. A scale parameter is associated to each fund in order to account for the behavioral specificity of the investors within funds. Note that \( \mu_l \) has been fixed to 1 because all the \( \mu_l \) are identifiable, except one.

### 6.3 The duration model

After the action, we model the duration until the next action. This duration is supposed to depend on this latter decision. We expect the investor not to change his mind when waiting for the next action. This assumption is made for easing the mathematical formulation of the likelihood function, and consequently the estimation of the combined model. In this section, we model the survival of the action \( \Lambda(c, t') \). The distribution of the observed duration between two actions is presented in Figure 11, proving the survival nature of the underlying phenomenon. A lifetime model has been chosen to handle the duration. The Weibull regression model has been retained, because it mimic the exponential model, but is more flexible.

\[ D(c, t') = \epsilon_{D,t'}, \text{ the mean of the random variable } \epsilon_{D,t'}. \] \( \epsilon_{D,t'} \) is assumed to follow a Weibull distribution. \( \lambda_D(c, t'|\theta, \omega_D, \beta) \) is the scale parameter of the Weibull distribution, and \( \eta_D \) the shape parameter.

\[ D(c, t') = \epsilon_{D,t'} \text{ with } \epsilon_{D,t'} \sim W(\lambda_D(c, t'|\theta, \omega_D, \beta), \eta_D). \] (25)

We define a utility

\[
m_D(c, t'|r_D, \theta, \beta) = \sum_{g \in \{1,2\}} I_{g,c} \text{ASC}_{D,R_D,g} + \sum_{g \in \{1,2\}} I_{g,c} \sum_{k=1}^{K_D-1} \theta_{D,k} \sum_{l=1}^{K_X} I_{D,g,k,l,R_D} X_{c,t'}(l) + \theta_{D,k_0} \text{Sigm} \nu_{X_{c,t'}}(l_0) I_{r=N} + \alpha_{D,N,1} I_{c,g=1} I_{r=N} D(c, t) + \alpha_{D,N,2} I_{c,g=2} I_{r=N} D(c, t) + \alpha_{D,R,1} I_{c,g=1} I_{r=R} D(c, t) + \theta_{B,R_D} V_B(c, t'|r_D, \beta) \] (26)

where \( I_{D,g,k,l,R_D} \) is an indicator equal to 1 if the parameter \( \theta_{D,k} \) is associated to the attribute \( X_{c,t}(l) \), to the group of funds \( g \), and associated to the risk \( r_D \). \( I_{g,c} \) is an
indicator equal to 1 if the company \( c \) belongs to the group of funds \( g \). \( I_{rD=N} \) is an indicator equal to 1 if \( r_D = N \), 0 otherwise. \( I_{rD=R} = 1 - I_{rD=N} \). In order get rid of the risk \( r_D \) in \( m_D(c, t' | r_D, \theta) \), we need to sum on levels of \( r_D \)

\[
m_D(c, t' | \theta, \omega_D, \beta) = m_D(c, t' | N, \theta, \beta) P_D(N, t' | \omega_D) + m_D(c, t' | R, \theta, \beta) P_D(N, t' | \omega_D),
\]

(27)

where \( P_D(N, t' | \omega_D) \) and \( P_D(R, t' | \omega_D) \) are shown in equation (16). The vector of parameters \( \theta \) is

\[
\theta = \{ ASC_{D,N}, ASC_{D,R}, \{ \theta_{D,k} \}_{k=1...K_D}, \alpha_{D,N,1}, \alpha_{D,N,2}, \alpha_{D,R,1}, \alpha_{D,R,2}, \{ \theta_{B,r_D} \}_{r_D=N,R} \}.
\]

(28)

We have

\[
\lambda_D(c, t' | \theta, \omega_D, \beta) = \frac{1}{e^{m_D(c, t' | \theta, \omega_D, \beta)}},
\]

(29)

consequently, if \( \Gamma() \) denotes the gamma function

\[
D(c, t') = \bar{e}_{D,t'} = \frac{1}{\lambda_D(c, t' | \theta, \omega_D)} \Gamma(1 + \frac{1}{\eta_D})
\]

\[
= e^{m_D(c, t' | \theta, \omega_D, \beta)} \Gamma(1 + \frac{1}{\eta_D}).
\]

(30)

The density of the Weibull distribution calculated for \( D(c, t') \) is

\[
f(D(c, t') | \lambda_D(c, t' | \theta, \omega_D, \beta), \eta_D) = \eta_D \lambda_D(c, t' | \theta, \omega_D, \beta) \eta_D D(c, t')^{\eta_D - 1} e^{-\lambda_D(c, t' | \theta, \omega_D, \beta) D(c, t')} \eta_D.
\]

(31)

In Equation (26), only attributes for \( t_H = 60 \) are used. They give the best model fit and provide interpretable parameters. Note that we did not present any descriptive statistics regarding \( D(c, t) \) in Section 5, because no generic correlation across funds appeared during the univariate analysis. In Equation (26), the influence of the previous duration \( D(c, t) \) for \( g = 1 \) and \( r_D = R \) has been discarded because the associated parameter did not appear to be significant (see Section 7).

6.4 The likelihood function

The likelihood of the action model is
\[
I_A(\beta, \omega_A) = \prod_{c=1}^{C} \prod_{t'=2}^{T} (P_{B}(c, t'|\beta, \omega_A)^{z_{B,c,t'}I_{c,t'}} \times P_{S}(c, t'|\beta, \omega_A)^{(1-z_{B,c,t'})I_{c,t'}},)
\]

C is the number of companies. \(z_{B,c,t'}\) is an indicator equal to 1 if stocks of c have been bought on \(t'\), 0 otherwise. \(I_{c,t'}\) is an indicator equal to 1 if an action has been observed on \(t'\) for c, 0 otherwise. Concerning the duration model, the likelihood function is

\[
I_D(\theta, \omega_D, \beta) = \prod_{c=1}^{C} \prod_{t'=1}^{T-1} f(D(c, t')|\lambda_D(c, t'|\theta, \omega_D, \beta), \eta_D)^{I_{c,t'}},
\]

where \(I_{c,t'}\) is an indicator equal to 1 if an action has been performed for c on \(t'\). Then, the joint likelihood function is

\[
I(\beta, \omega_A, \theta, \omega_D) = I_A(\beta, \omega_A)I_D(\theta, \omega_D, \beta)
\]

and the log-likelihood function

\[
L(\beta, \omega_A, \theta, \omega_D) = \log(I(\beta, \omega_A, \theta, \omega_D))
\]

7 Model estimation

The combined model is estimated by maximum likelihood using the biogeme software (Bierlaire, 2003 and Bierlaire and Fétiarison, 2009). The log-likelihood for the entire model is presented in equation (35). The processed data were used for estimation (see Sections 4 and 5). General estimation results are displayed in Table 5. The 61 parameters of the combined model are split between the action and duration models (respectively 29 and 32 parameters). The number of observations for the duration model is equal to the number of actions (9178), minus the number of companies (1121). For the last observed decision associated to a company, we do not know the duration until the next action. The duration model has a big impact on the log-likelihood of the combined model. The \(R^2\) of the duration model is low. Interpretations of the parameters are explained in the following for the action and duration models.

7.1 The action model

Parameters values and associated t-tests are presented in Tables 11 and 12.

- the risk perception in the action model: ASC\(_{WA}\) was not significantly different from minus the VIX threshold defined in Figure 3. This value was the starting value for estimation. \(\omega_{A,1}\) and \(\omega_{A,2}\) are the two parameters associated
<table>
<thead>
<tr>
<th></th>
<th>Action choice model</th>
<th>Duration model</th>
<th>Combined model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nb parameters</td>
<td>29</td>
<td>32</td>
<td>61</td>
</tr>
<tr>
<td>Nb observations</td>
<td>9178</td>
<td>8057</td>
<td>9178</td>
</tr>
<tr>
<td>Null Log-likelihood</td>
<td>-6361.705</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final Log-likelihood</td>
<td>-5635.580</td>
<td>-19711.528</td>
<td>-25347.109</td>
</tr>
<tr>
<td>$\bar{p}^2/R^2$</td>
<td>0.109</td>
<td>0.048</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: General estimation results of the action and duration models

to the VIX, respectively for the groups of funds 1 and 2. Logically, both parameters are positive meaning that the risk increases with the VIX, as described in Sections 6.1.1 and 6.1.2. $\omega_{A,1} > \omega_{A,2}$, so investors managing funds in group 1 are more sensitive to risk than investors managing group 2. $\omega_{A,3}$ is associated to the standard error of sentiment calculated for $t_H = 5 \text{ (days)}$. It is positive as expected, showing that the increase of the fluctuations in the analyst opinions increases the perception of risk, which is logical.

- The alternative specific constants: $ASC_{B,N}$ and $ASC_{R,N}$ are negative, meaning that in the two risk situations, buy is penalized.

- The parameters associated with the explanatory variables: A statistical analysis has been conducted (see Table 3) and several model specifications have been tested. The variables which appeared to be the best appropriated to this analysis, were calculated with $t_H = 1 \text{ (day)}$, meaning that the investors base their choices on short-term information. No $\beta_{B,k}$ ($k = 1 \ldots 15$) is generic across risk situations and groups of funds. The generic parameters across groups of funds are $\beta_{B,4}$, $\beta_{B,5}$, $\beta_{B,6}$ and $\beta_{B,15}$. $\beta_{B,6}$ and $\beta_{B,15}$ are positive. Under the risk N, the increase of the short-term values of quality and value favor the buy alternative. $\beta_{B,5}$ is negative, an increase of the long-term value of quality favor the sell alternative. $\beta_{B,4}$ is positive, so under the risk R, the increase of the short-term value of price increases the probability to buy, underlying the tendency of investors for taking advantage of immediate fluctuations of variables.

- The memory effect: $\alpha_{B,N}$ and $\alpha_{B,R}$ are both negative showing that for a given stock, investors have not the tendency to perform consecutively two buy actions. This is stronger in under the risk R than in under the risk N. Investors are more likely to bet on short-term returns in under R, which is logical. As expected, this effect is attenuated with the increase of the duration between two consecutive actions, as shown by the negative value of $\lambda_{B,N}$ and $\lambda_{B,R}$. The attenuation is higher in a normal situation than in a risky situation.
\textbf{7.2 The duration model}

Parameters values and associated t-tests are shown and in Table 13 and 14.

- **The risk perception in the duration model:** \(\text{ASC}_{W_d,1}\) and \(\text{ASC}_{W_d,2}\) are negative, as expected. Investors of fund groups 1 have a stronger a priori toward the risk \(N\), compared to investors of group 2 (\(\text{ASC}_{W_d,1} < \text{ASC}_{W_d,2}\)). \(\omega_{D,1}\) and \(\omega_{D,2}\) are positive as expected. Investors of fund group 1 are more sensitive to risk than investors of group 2 (\(\omega_{D,1} > \omega_{D,2}\)).

- **The shape parameter:** \(\eta_D\) is the shape parameter of the Weibull regression model, which has been estimated under 1, showing the proximity to a an exponential regression model, but with a distribution characterized by a less heavy right tail.

- **The constants:** \(\text{ASC}_{D,N,1}\), \(\text{ASC}_{D,N,2}\), \(\text{ASC}_{D,R,1}\), \(\text{ASC}_{D,R,2}\) are all positive. They characterize the average duration in pure risk situations and are specific to the fund groups, all other attributes being equal to 0. The average of the predicted duration is presented in Equation (30). The average of the duration in the situation \(N\) for the fund group 1 is 68.9 weeks, and for group 2, 57.6 weeks. The average of the duration under risk \(R\) for the fund group 1 is 2.5 weeks, and for group 2, 2.9 weeks. This seems logical, because investors have to react much more faster in risky situations than in normal situations.

- **The parameters associated with the explanatory variables:** No parameter is generic across fund groups and risk situations, which underline the strong specificities of the investors’ behaviors. \(\theta_{D,16}\) is associated to the standard error of the VIX calculated over 360 days, for the fund group 2 under the risk \(N\). It is negative, which is logical because when the variation of the principle risk indicator of the market increases, investors have tendency to perform actions more often. \(\theta_{B,N}\) and \(\theta_{B,R}\) are negative, so when the utility of the buy alternative increases, the
duration decreases under both risk N and R. In any case, when money is invested in a company, the vigilance of the investor toward the company increases, and he is more likely to adjust his decision in a near future.

- **The influence of the previous duration**: $\alpha_{D,N,1}$, $\alpha_{D,R,1}$ and $\alpha_{D,N,2}$ are positive. This shows the stability of the phenomenon, in the sense that the increase of the previous duration generates an increase of the current duration.

To conclude this section, the parameters of the combined model are significant and interpretable. The interpretations have been discussed with the involved investors. The variables presented in the Section 5 are adapted to explain the investors’ behavior. Behavioral Specificities within each fund appear clearly.

8 Model prediction

In this section we study the prediction accuracy of the combined model. We start by examining the model prediction on the estimation data and perform a cross-validation. For the action model, we conduct a simulation analysis.

8.1 Prediction on the estimation data

The frequencies of the predicted probabilities of the observed actions are shown in Figure 12. If the model was perfect, all predicted probabilities should be equal to 1. This corresponds to a log-likelihood equal to 0. This is of course never the case, but a significant shift of the distribution on the right is observed. In addition 66.25% of the actions are predicted with a probability higher than 0.5, represented by the grey bin. The model is compared to a simple binary logit model. It contains only two parameters which are the constants in the deterministic utilities of *buy* and *sell*, and without risk perception. It has the property to reproduce the aggregated shares of actions of the estimation data, when used for prediction. In that case, there are approximately as much actions of type *Buy* (4530) than action of type *Sell* (4648). This simple model predicts a quasi equal probability for the two actions. Compared to this simple model, the proposed model increases the prediction accuracy of 16.25%.

The standardized residual of the duration model $z_{c,t}$ is defined as

$$z_{c,t} = \eta_D(\log(D(c, t)) - m_D(c, t|\theta, \omega_D, \beta)) \sim EV(0, 1),$$

which makes the parallel with the normal regression where the standardized residuals of the logarithm of the dependent variable are supposed normally distributed $N(0, 1)$. The distribution of $\{z_{c,t}\}$ is plotted in Figure 13 (histogram), as well as the theoretical distribution $EV(0, 1)$ (curve). The observed distribution is near from the theoretical
curve, but the model tends to over-predict the duration for some observations. A bimodality appears. The over-predictions concerns very small durations. The duration model works well for situations presented in Figure 4 where actions are sparse, but not for situations displayed in Figure 5, where actions are concentrated. Several regression models have been tested, such as the lognormal, Poisson, negative binomial, exponential and Rayleigh (these two latter are particular cases of the Weibull). In terms of fit and residual analysis, the Weibull appeared to be the best. Regarding the specification, many trials have been done. The improvement of the residual distribution is possible by the using of distribution mixtures and the integration of supplementary data, such as money flows.

This prediction analysis is performed on the estimation data. It reinforces the estimation results (see Section 7), but the models have not been yet tested for forecasting.

8.2 Cross-validation

We need to check the prediction capability of the combined model. The cross-validation consists in estimating the combined model on a part of the data, and simulate on the remaining part. The total time horizon is divided in five periods of equal duration. The starting date and ending date of each period as well as the number of actions per period are shown in Table 6. The estimation of the combined model is done on 4 subsets and the simulation on the remaining subset. The experience is repeated five times in order to cover all the possibilities.

<table>
<thead>
<tr>
<th>Validation set</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starting date</td>
<td>2005.01.03</td>
<td>2006.03.02</td>
<td>2007.04.25</td>
<td>2008.06.05</td>
<td>2009.07.24</td>
</tr>
<tr>
<td>Ending date</td>
<td>2006.03.01</td>
<td>2007.04.24</td>
<td>2008.06.04</td>
<td>2009.07.23</td>
<td>2010.09.13</td>
</tr>
<tr>
<td>Nb actions</td>
<td>1363</td>
<td>1511</td>
<td>1766</td>
<td>2149</td>
<td>2399</td>
</tr>
</tbody>
</table>

Table 6: Starting dates, ending dates and number of actions per subset of data for the cross-validation

For each experience, we calculate statistics revealing the prediction accuracy of the models on the simulation subset. Concerning the action model, the predicted log-likelihood is calculated (Predicted $\mathcal{L}$). For the duration model, $R^2_c$ is the predicted $R^2$ for experience c. These values are respectively compared to the log-likelihood (Estimated $\mathcal{L}$) and the $R^2$ (Estimated $R^2$) obtained when applying the combined model (estimated on the entire data) on the simulation subset.
The results are presented in Table 7. Logically the estimated $R^2$ and $\mathcal{L}$ are always higher than $R^2_0$ and the predicted $\mathcal{L}$. Regarding the action model, the log-likelihood increases chronologically because the volume of decisions is also increasing (see Table 6). The higher difference between the estimated and predicted log-likelihood is observed for the experience 5 (235.141). However for every experiences, both log-likelihoods have the same order. This shows the stability of the action model. This is not the case for the duration model. We explicit $R^2_c$, $t_{0,c}$ is the starting date of the simulation subset for experience $c$, and $T_c$ is the ending date. $D(c, t)$ is the observed duration for day $t$ and company $c$, $\hat{D}(c, t)$ is the predicted duration, and $\bar{D}(c, t)$ is the duration mean calculated over $t$ and $c$.

$$R^2_c = 1 - \frac{\sum_{c=1}^{c} \sum_{t=t_{0,c}}^{T_c} (\hat{D}(c, t) - D(c, t))^2}{\sum_{c=1}^{c} \sum_{t=t_{0,c}}^{T_c} (D(c, t) - \bar{D}(c, t))^2} \tag{38}$$

$R^2_c$ is negative for $c \in \{3, 4, 5\}$. The predictions are worse than those of the simple model predicting the duration mean on the considered time period. The experience 5 is even not well predicted by the duration model estimated on the entire data, further investigations have to be performed.

<table>
<thead>
<tr>
<th>Experience</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2_c$</td>
<td>0.034</td>
<td>0.047</td>
<td>-0.003</td>
<td>-0.003</td>
<td>-1.790</td>
</tr>
<tr>
<td>Estimated $R^2$</td>
<td>0.053</td>
<td>0.065</td>
<td>0.010</td>
<td>0.025</td>
<td>-0.464</td>
</tr>
<tr>
<td>Predicted $\mathcal{L}$</td>
<td>-394.436</td>
<td>-431.504</td>
<td>-1095.921</td>
<td>-1269.583</td>
<td>-1588.811</td>
</tr>
<tr>
<td>Estimated $\mathcal{L}$</td>
<td>-383.671</td>
<td>-420.055</td>
<td>-1052.993</td>
<td>-1220.002</td>
<td>-1353.67</td>
</tr>
</tbody>
</table>

Table 7: Results of the cross-validation performed on the estimation data

Specifically to the action model and for each experience, we have calculated the percentage of observations predicted with a probability less than 0.5, which are considered badly predicted. The results are displayed in Table 8. For the action model estimated on the entire data, there are 33.75% of bad predictions. For every experience, the percentage is similar to this value. This underlines the stability of this model.

<table>
<thead>
<tr>
<th>Experience</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action model</td>
<td>36.90</td>
<td>38.65</td>
<td>31.65</td>
<td>30.15</td>
<td>41.31</td>
</tr>
</tbody>
</table>

Table 8: Percentages of badly predicted observations per experience of cross-validation

The financial crisis of 2008 appears in the subset 4 (see Table 6). In this period, the prediction of the duration model are bad ($R^2_d < 0$). This is not the case for the action model (30.15% of badly predicted observation), which shows its robustness. The cross-validation is a first step toward forecasting, the results are worth for the action model and limited for the duration model.
8.3 Simulation

In this section, we present a concrete forecasting application of the action model. The experience 5 of the cross-validation is considered (see Table 6). The action model is estimated on the calibration subset and applied on the simulation subset. We hypothesize that the action days are fixed. Five simulations are performed on the simulation subset. In each simulation and for each action day, an action is drawn from the predicted probability distribution. The number of buy and sell actions are aggregated per month. Results are presented in Table 9. For each month, the number of buy actions, nb\_buy\_sim, and sell actions nb\_sell\_sim, are shown as “nb\_buy\_sim/nb\_sell\_sim”. The observed shares are also displayed in the column “Reality” (“nb\_buy\_obs/nb\_sell\_obs”).

<table>
<thead>
<tr>
<th>Month</th>
<th>Reality</th>
<th>Simul. 1</th>
<th>Simul. 2</th>
<th>Simul. 3</th>
<th>Simul. 4</th>
<th>Simul. 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jul. 09</td>
<td>21/31</td>
<td>20/32</td>
<td>25/27</td>
<td>26/26</td>
<td>22/30</td>
<td>18/34</td>
</tr>
<tr>
<td>Sep. 09</td>
<td>108/84</td>
<td>91/101</td>
<td>106/86</td>
<td>116/76</td>
<td>101/91</td>
<td>108/84</td>
</tr>
<tr>
<td>Oct. 09</td>
<td>110/141</td>
<td>121/130</td>
<td>130/121</td>
<td>116/135</td>
<td>116/135</td>
<td>112/139</td>
</tr>
<tr>
<td>Nov. 09</td>
<td>105/138</td>
<td>109/134</td>
<td>112/131</td>
<td>117/126</td>
<td>113/130</td>
<td>104/139</td>
</tr>
<tr>
<td>Dec. 09</td>
<td>76/93</td>
<td>75/94</td>
<td>86/83</td>
<td>77/92</td>
<td>75/94</td>
<td>58/111</td>
</tr>
<tr>
<td>Jan. 10</td>
<td>69/82</td>
<td>71/80</td>
<td>70/81</td>
<td>67/84</td>
<td>69/82</td>
<td>70/81</td>
</tr>
<tr>
<td>Feb. 10</td>
<td>69/77</td>
<td>68/78</td>
<td>61/85</td>
<td>56/90</td>
<td>68/78</td>
<td>67/79</td>
</tr>
<tr>
<td>Mar. 10</td>
<td>101/79</td>
<td>87/93</td>
<td>90/90</td>
<td>92/88</td>
<td>94/86</td>
<td>91/89</td>
</tr>
<tr>
<td>Apr. 10</td>
<td>148/54</td>
<td>96/106</td>
<td>112/90</td>
<td>104/98</td>
<td>101/101</td>
<td>108/94</td>
</tr>
<tr>
<td>May 10</td>
<td>85/84</td>
<td>76/93</td>
<td>85/84</td>
<td>75/94</td>
<td>83/86</td>
<td>86/83</td>
</tr>
<tr>
<td>Jun. 10</td>
<td>41/48</td>
<td>34/55</td>
<td>38/51</td>
<td>41/48</td>
<td>49/40</td>
<td>40/49</td>
</tr>
<tr>
<td>Jul. 10</td>
<td>53/96</td>
<td>65/84</td>
<td>77/72</td>
<td>78/71</td>
<td>72/77</td>
<td>81/68</td>
</tr>
<tr>
<td>Aug. 10</td>
<td>79/93</td>
<td>80/92</td>
<td>86/86</td>
<td>64/108</td>
<td>86/86</td>
<td>74/98</td>
</tr>
<tr>
<td>Sep. 10</td>
<td>17/15</td>
<td>23/9</td>
<td>14/18</td>
<td>16/16</td>
<td>20/12</td>
<td>15/17</td>
</tr>
</tbody>
</table>

Table 9: Results of the simulations performed with the action model on the period going from 2009.07.24 to 2010.09.13

For each month, the simulated and observed shares are compared. We define the percentage of error

\[ Err = \frac{\text{nb\_buy\_obs} - \text{nb\_buy\_sim}}{\text{nb\_buy\_obs} + \text{nb\_sell\_obs}} \times 100, \]  

(39)

it corresponds to the number of false simulated actions divided by the total number
of actions within the month. As the action days are fixed, we have \( n_{\text{buy obs}} + n_{\text{sell obs}} = n_{\text{buy sim}} + n_{\text{sell sim}} \). The percentages of error are shown in Table 10. The highest values are observed for April 2010, otherwise no percentage of error is above 20\%. Moreover, 81.33\% of the percentages of error are under 10\%, showing the forecasting accuracy of the action model. The aggregation of the results has been done per month, which is relatively detailed specially for long-term investments. This simulation emphasizes the good quality and usefulness of the model in real-life applications.

<table>
<thead>
<tr>
<th>Month</th>
<th>Simul. 1</th>
<th>Simul. 2</th>
<th>Simul. 3</th>
<th>Simul. 4</th>
<th>Simul. 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jul. 09</td>
<td>1.92</td>
<td>-7.69</td>
<td>-9.62</td>
<td>-1.92</td>
<td>5.77</td>
</tr>
<tr>
<td>Aug. 09</td>
<td>5.94</td>
<td>12.38</td>
<td>12.38</td>
<td>13.37</td>
<td>6.44</td>
</tr>
<tr>
<td>Sep. 09</td>
<td>8.85</td>
<td>1.04</td>
<td>-4.17</td>
<td>3.65</td>
<td>0.00</td>
</tr>
<tr>
<td>Oct. 09</td>
<td>-4.38</td>
<td>-7.97</td>
<td>-2.39</td>
<td>-2.39</td>
<td>-0.80</td>
</tr>
<tr>
<td>Nov. 09</td>
<td>-1.65</td>
<td>-2.88</td>
<td>-4.94</td>
<td>-3.29</td>
<td>0.41</td>
</tr>
<tr>
<td>Dec. 09</td>
<td>0.59</td>
<td>-5.92</td>
<td>-0.59</td>
<td>0.59</td>
<td>10.65</td>
</tr>
<tr>
<td>Jan. 10</td>
<td>-1.32</td>
<td>-0.66</td>
<td>1.32</td>
<td>0.00</td>
<td>-0.66</td>
</tr>
<tr>
<td>Feb. 10</td>
<td>0.68</td>
<td>5.48</td>
<td>8.90</td>
<td>0.68</td>
<td>1.37</td>
</tr>
<tr>
<td>Mar. 10</td>
<td>7.78</td>
<td>6.11</td>
<td>5.00</td>
<td>3.89</td>
<td>5.56</td>
</tr>
<tr>
<td>Apr. 10</td>
<td>25.74</td>
<td>17.82</td>
<td>21.78</td>
<td>23.27</td>
<td>19.80</td>
</tr>
<tr>
<td>May 10</td>
<td>5.33</td>
<td>0.00</td>
<td>5.92</td>
<td>1.18</td>
<td>-0.59</td>
</tr>
<tr>
<td>Jun. 10</td>
<td>7.87</td>
<td>3.37</td>
<td>0.00</td>
<td>-8.99</td>
<td>1.12</td>
</tr>
<tr>
<td>Jul. 10</td>
<td>-8.05</td>
<td>-16.11</td>
<td>-16.78</td>
<td>-12.75</td>
<td>-18.79</td>
</tr>
<tr>
<td>Aug. 10</td>
<td>-0.58</td>
<td>-4.07</td>
<td>8.72</td>
<td>-4.07</td>
<td>2.91</td>
</tr>
</tbody>
</table>

Table 10: Percentages of error (see Equation (39)) based on simulations of Table 9

Note that a simulation using the combined model has been tested. The obtained results were not convincing. The reason is the limited predictive power of the duration model (see Section 8.2). Consequently we focused on the action model.

9 Conclusion

We developed models capturing the investors' behavior. The data provided by the Swiss bank Lombard Odier Darier Hentsch have been processed to infer the deci-
sions performed by investors. Variables are computed based on the five indicators and the VIX, for reflecting the dynamics of the observed behavior. A correlation analysis allows to understand the links between the decisions and the decisional context. An integrated approach has been proposed to model simultaneously the choice of action (buy or sell) and the duration between two actions. A binary logit model with latent classes is developed for the choice of action, where the latent classes correspond to the risk situations. Two situations are considered, normal and risky. The duration is handled using a Weibull regression, which also accounts for the two risk situations. But, the perception is different compared to the action model. Both models account explicitly for the dynamics of the behavioral phenomenon. The two models are linked, because the perception of the buy alternative enters in the duration model.

The combination of the action and duration models is estimated simultaneously, using the processed data. Parameters are interpretable. Results have been discussed with investors. They explicit causalities and reveal behavioral mechanisms. The accounting of the dynamics has sense, investors consider their previous decisions when taking current decisions. The specificity of the investors' behavior within each fund appears. The hypothesis about the risk perception is valid in the action and duration models. In risky situations, the duration between two consecutive actions is shorter than in normal situations, which is logical, as investors tend to adjust more often their portfolio in risky situations. Regarding the action choice model, the specificity of the behavior per fund is more important in risky than in normal situations. This is logical, as individual personalities are emphasized in panic situations.

Predictions of the models have been checked and the combined model has been cross-validated on the estimation data. The predictive accuracy of the action model is good, and limited for the duration model. This is not surprising for the duration model because the associated modeling assumptions are moved from the reality. This is due to the fact that we have decided to develop the most realistic models remaining operational. In addition, crucial data are missing about money inflows and outflows to characterize this duration. The relevance of the action model has been shown by the simulation which is the practical way to use it. As it is, the action model can be embedded in a simulator for forecasting aggregated action shares based on market scenarios. For investors, this is a relevant decision-aid tool.

There are several perspectives to this work. In terms of modeling, the risk perception can be refined for both models. More than two latent classes could be considered. The integration of relevant supplementary data can also be considered, such as portfolio information, money inflows and outflows, and investors' characteristics. Dedicated data collection could be conducted in order to point out precisely the information used by investors when taking decisions and refine the explanatory variables. Then, the proposed model can be improved, and the priority. should be put on the duration model.
For this latter, mixtures of distribution can be considered. In addition, the assumption about the non revision of the duration once the decision has been taken, can be relaxed. The forecasting accuracy of the action model has been tested by simulation. The same thing with the combined model could be considered with a better duration model.

Regarding the evolution of the stock prices, the developed model is not sufficient, other financial actors should be modeled, in order to capture the entire financial scene. This requires to have specific and detailed behavioral data about all the actors and data about the scene. Then the different models could be embedded into a single simulator, allowing the prediction of the evolution of stock prices.
Figure 11: Distribution of the duration expressed in weeks (5 days)
Figure 12: Distribution of the predicted action choice probabilities calculated on the estimation data
Figure 13: Comparison between the distribution of the residuals \( \{z_{c,t}\} \), related to the duration model (histogram) and the theoretical distribution (curve)
References


URL: http://www.sciencedirect.com/science/article/B6VC0-494HPGM-1/2/aca7702ca7b7e4807d312fd243c55f1e


39


URL: [http://www.sciencedirect.com/science/article/B6VC0-3VW1T8-6/2/2ce680f1a1dc9b5fe1e056e3c0faa2d](http://www.sciencedirect.com/science/article/B6VC0-3VW1T8-6/2/2ce680f1a1dc9b5fe1e056e3c0faa2d)


URL: [http://www.sciencedirect.com/science/article/B6V85-419JHMW-4/2/e00f233b151a01fd8c43e8574e39a524](http://www.sciencedirect.com/science/article/B6V85-419JHMW-4/2/e00f233b151a01fd8c43e8574e39a524)


URL:  http://www.sciencedirect.com/science/article/B6V33-4DB4VS8-2/2/4d6e679fee1b7982e7ab818c786e51af


### Table 11: Estimated parameters of the action model (β)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Transform</th>
<th>$t_H$ (day)</th>
<th>$g$</th>
<th>$r$</th>
<th>Value</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC$_{B,N}$</td>
<td></td>
<td>1,2</td>
<td>N</td>
<td>-1.39</td>
<td>4.41</td>
<td></td>
</tr>
<tr>
<td>ASC$_{B,R}$</td>
<td></td>
<td>1,2</td>
<td>R</td>
<td>-2.01</td>
<td>4.44</td>
<td></td>
</tr>
<tr>
<td>$\beta_{B,1}$</td>
<td>Price</td>
<td>Long</td>
<td>1</td>
<td>N</td>
<td>-2.31</td>
<td>3.15</td>
</tr>
<tr>
<td>$\beta_{B,2}$</td>
<td>Price</td>
<td>Perf</td>
<td>1</td>
<td>2</td>
<td>N,R</td>
<td>6.18</td>
</tr>
<tr>
<td>$\beta_{B,3}$</td>
<td>Price</td>
<td>Short</td>
<td>1</td>
<td>2</td>
<td>N</td>
<td>-1.02</td>
</tr>
<tr>
<td>$\beta_{B,4}$</td>
<td>Price</td>
<td>Short</td>
<td>1</td>
<td>1,2</td>
<td>R</td>
<td>0.916</td>
</tr>
<tr>
<td>$\beta_{B,5}$</td>
<td>Quality</td>
<td>Long</td>
<td>1</td>
<td>1,2</td>
<td>N</td>
<td>-0.428</td>
</tr>
<tr>
<td>$\beta_{B,6}$</td>
<td>Quality</td>
<td>Short</td>
<td>1</td>
<td>1,2</td>
<td>N</td>
<td>0.723</td>
</tr>
<tr>
<td>$\beta_{B,7}$</td>
<td>Sentiment</td>
<td>Long</td>
<td>1</td>
<td>2</td>
<td>N</td>
<td>0.659</td>
</tr>
<tr>
<td>$\beta_{B,8}$</td>
<td>Technic</td>
<td>Long</td>
<td>1</td>
<td>1</td>
<td>R</td>
<td>1.06</td>
</tr>
<tr>
<td>$\beta_{B,9}$</td>
<td>Technic</td>
<td>Long</td>
<td>1</td>
<td>1</td>
<td>N</td>
<td>1.02</td>
</tr>
<tr>
<td>$\beta_{B,10}$</td>
<td>Technic</td>
<td>Long</td>
<td>1</td>
<td>2</td>
<td>R</td>
<td>-0.805</td>
</tr>
<tr>
<td>$\beta_{B,11}$</td>
<td>Technic</td>
<td>Short</td>
<td>1</td>
<td>1</td>
<td>R</td>
<td>1.80</td>
</tr>
<tr>
<td>$\beta_{B,12}$</td>
<td>Technic</td>
<td>Short</td>
<td>1</td>
<td>1</td>
<td>N</td>
<td>2.78</td>
</tr>
<tr>
<td>$\beta_{B,13}$</td>
<td>Technic</td>
<td>Short</td>
<td>1</td>
<td>2</td>
<td>R</td>
<td>1.30</td>
</tr>
<tr>
<td>$\beta_{B,14}$</td>
<td>Technic</td>
<td>Short</td>
<td>1</td>
<td>2</td>
<td>N</td>
<td>-1.47</td>
</tr>
<tr>
<td>$\beta_{B,15}$</td>
<td>Value</td>
<td>Short</td>
<td>1</td>
<td>1,2</td>
<td>N</td>
<td>0.625</td>
</tr>
<tr>
<td>$\alpha_N$</td>
<td></td>
<td>1,2</td>
<td>N</td>
<td>-0.222</td>
<td>5.43</td>
<td></td>
</tr>
<tr>
<td>$\alpha_R$</td>
<td></td>
<td>1,2</td>
<td>R</td>
<td>-0.312</td>
<td>6.77</td>
<td></td>
</tr>
<tr>
<td>$\lambda_{B,N}$</td>
<td></td>
<td>1,2</td>
<td>N</td>
<td>-0.0156</td>
<td>3.79</td>
<td></td>
</tr>
<tr>
<td>$\lambda_{B,R}$</td>
<td></td>
<td>1,2</td>
<td>R</td>
<td>-0.00946</td>
<td>5.27</td>
<td></td>
</tr>
<tr>
<td>$\mu_1$</td>
<td></td>
<td>1,2</td>
<td>N,R</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_2$</td>
<td></td>
<td>1,2</td>
<td>N,R</td>
<td>1.23</td>
<td>5.07</td>
<td></td>
</tr>
<tr>
<td>$\mu_3$</td>
<td></td>
<td>1,2</td>
<td>N,R</td>
<td>1.58</td>
<td>4.21</td>
<td></td>
</tr>
<tr>
<td>$\mu_4$</td>
<td></td>
<td>1,2</td>
<td>N,R</td>
<td>2.93</td>
<td>3.16</td>
<td></td>
</tr>
<tr>
<td>$\mu_5$</td>
<td></td>
<td>1,2</td>
<td>N,R</td>
<td>1.43</td>
<td>2.77</td>
<td></td>
</tr>
<tr>
<td>$\mu_6$</td>
<td></td>
<td>1,2</td>
<td>N,R</td>
<td>2.46</td>
<td>3.48</td>
<td></td>
</tr>
<tr>
<td>Parameter</td>
<td>$\nu$</td>
<td>Transform</td>
<td>$t_H$ (day)</td>
<td>$g$</td>
<td>Value</td>
<td>t-test</td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>-----------</td>
<td>-------------</td>
<td>-----</td>
<td>-------</td>
<td>--------</td>
</tr>
<tr>
<td>ASC_{W,1}</td>
<td>1</td>
<td></td>
<td>1,2</td>
<td>-25.327</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega_{A,1}$</td>
<td>VIX</td>
<td></td>
<td>1</td>
<td>1.37</td>
<td>49.41</td>
<td></td>
</tr>
<tr>
<td>$\omega_{A,2}$</td>
<td>VIX</td>
<td></td>
<td>2</td>
<td>1.08</td>
<td>31.39</td>
<td></td>
</tr>
<tr>
<td>$\omega_{A,3}$</td>
<td>Sentiment</td>
<td>Sigm</td>
<td>5</td>
<td>1,2</td>
<td>9.29</td>
<td>5.17</td>
</tr>
</tbody>
</table>

Table 12: Estimated parameters of the risk model associated to the action model ($\omega_A$)
<table>
<thead>
<tr>
<th>Parameter</th>
<th>( \nu )</th>
<th>Transform</th>
<th>( t_H ) (day)</th>
<th>g</th>
<th>r</th>
<th>Value</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC(_{D,N,1})</td>
<td>1</td>
<td></td>
<td>1</td>
<td>N</td>
<td></td>
<td>3.54</td>
<td>5.49</td>
</tr>
<tr>
<td>ASC(_{D,N,2})</td>
<td>1</td>
<td></td>
<td>2</td>
<td>N</td>
<td></td>
<td>3.36</td>
<td>10.71</td>
</tr>
<tr>
<td>ASC(_{D,R,1})</td>
<td>1</td>
<td></td>
<td>1</td>
<td>N</td>
<td></td>
<td>0.242</td>
<td>1.54</td>
</tr>
<tr>
<td>ASC(_{D,R,2})</td>
<td>1</td>
<td></td>
<td>2</td>
<td>N</td>
<td></td>
<td>0.398</td>
<td>2.93</td>
</tr>
<tr>
<td>( \theta_{D,1} )</td>
<td>Price</td>
<td>Short()</td>
<td>60</td>
<td>1</td>
<td>N</td>
<td>-1.22</td>
<td>-3.47</td>
</tr>
<tr>
<td>( \theta_{D,2} )</td>
<td>Price</td>
<td>Sigm()</td>
<td>60</td>
<td>1</td>
<td>N</td>
<td>1.52</td>
<td>1.78</td>
</tr>
<tr>
<td>( \theta_{D,3} )</td>
<td>Quality</td>
<td>Long()</td>
<td>60</td>
<td>1</td>
<td>N</td>
<td>-1.08</td>
<td>-3.03</td>
</tr>
<tr>
<td>( \theta_{D,4} )</td>
<td>Quality</td>
<td>Short()</td>
<td>60</td>
<td>1</td>
<td>R</td>
<td>0.960</td>
<td>1.80</td>
</tr>
<tr>
<td>( \theta_{D,5} )</td>
<td>Quality</td>
<td>Short()</td>
<td>60</td>
<td>2</td>
<td>R</td>
<td>-0.661</td>
<td>-1.54</td>
</tr>
<tr>
<td>( \theta_{D,6} )</td>
<td>Quality</td>
<td>Short()</td>
<td>60</td>
<td>1</td>
<td>N</td>
<td>-1.43</td>
<td>-1.96</td>
</tr>
<tr>
<td>( \theta_{D,7} )</td>
<td>Sentiment</td>
<td>Long()</td>
<td>60</td>
<td>2</td>
<td>R</td>
<td>-0.716</td>
<td>-3.29</td>
</tr>
<tr>
<td>( \theta_{D,8} )</td>
<td>Sentiment</td>
<td>Short()</td>
<td>60</td>
<td>2</td>
<td>R</td>
<td>0.990</td>
<td>3.72</td>
</tr>
<tr>
<td>( \theta_{D,9} )</td>
<td>Technic</td>
<td>Long()</td>
<td>60</td>
<td>2</td>
<td>R</td>
<td>1.42</td>
<td>5.38</td>
</tr>
<tr>
<td>( \theta_{D,10} )</td>
<td>Technic</td>
<td>Long()</td>
<td>60</td>
<td>1</td>
<td>N</td>
<td>-1.18</td>
<td>-2.36</td>
</tr>
<tr>
<td>( \theta_{D,11} )</td>
<td>Technic</td>
<td>Short()</td>
<td>60</td>
<td>1</td>
<td>R</td>
<td>1.79</td>
<td>3.19</td>
</tr>
<tr>
<td>( \theta_{D,12} )</td>
<td>Technic</td>
<td>Sigm()</td>
<td>60</td>
<td>2</td>
<td>N</td>
<td>-1.48</td>
<td>-3.91</td>
</tr>
<tr>
<td>( \theta_{D,13} )</td>
<td>Value</td>
<td>Short()</td>
<td>60</td>
<td>2</td>
<td>R</td>
<td>1.90</td>
<td>4.84</td>
</tr>
<tr>
<td>( \theta_{D,14} )</td>
<td>Value</td>
<td>Short()</td>
<td>60</td>
<td>1</td>
<td>N</td>
<td>2.25</td>
<td>2.60</td>
</tr>
<tr>
<td>( \theta_{D,15} )</td>
<td>Value</td>
<td>Sigm()</td>
<td>60</td>
<td>2</td>
<td>N</td>
<td>-0.613</td>
<td>-1.72</td>
</tr>
<tr>
<td>( \theta_{D,16} )</td>
<td>VIX</td>
<td>Sigm()</td>
<td>360</td>
<td>2</td>
<td>N</td>
<td>-2.05</td>
<td>-4.99</td>
</tr>
<tr>
<td>( \theta_{B,N} )</td>
<td></td>
<td></td>
<td>1,2</td>
<td>N</td>
<td></td>
<td>-0.350</td>
<td>-2.27</td>
</tr>
<tr>
<td>( \theta_{B,R} )</td>
<td></td>
<td></td>
<td>1,2</td>
<td>R</td>
<td></td>
<td>-0.261</td>
<td>-2.53</td>
</tr>
<tr>
<td>( \alpha_{D,N,1} )</td>
<td></td>
<td></td>
<td>1</td>
<td>N</td>
<td></td>
<td>7.09</td>
<td>4.77</td>
</tr>
<tr>
<td>( \alpha_{D,R,1} )</td>
<td></td>
<td></td>
<td>1</td>
<td>R</td>
<td></td>
<td>5.29</td>
<td>3.25</td>
</tr>
<tr>
<td>( \alpha_{D,N,2} )</td>
<td></td>
<td></td>
<td>2</td>
<td>R</td>
<td></td>
<td>2.23</td>
<td>1.84</td>
</tr>
<tr>
<td>( \eta_D )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.530</td>
<td>82.26</td>
</tr>
</tbody>
</table>

Table 13: Estimated parameters of the duration model (\( \theta \))

45
<table>
<thead>
<tr>
<th>Parameter</th>
<th>ν</th>
<th>Transform</th>
<th>tH (day)</th>
<th>g</th>
<th>Value</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASCW₀,₁</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>-7.48</td>
<td>-1.59</td>
</tr>
<tr>
<td>ASCW₀,₂</td>
<td>1</td>
<td></td>
<td></td>
<td>2</td>
<td>-4.93</td>
<td>-3.76</td>
</tr>
<tr>
<td>ωD,₁</td>
<td>VIX</td>
<td></td>
<td></td>
<td>1</td>
<td>0.377</td>
<td>2.10</td>
</tr>
<tr>
<td>ωD,₂</td>
<td>VIX</td>
<td></td>
<td></td>
<td>2</td>
<td>0.263</td>
<td>4.53</td>
</tr>
</tbody>
</table>

Table 14: Estimated parameters of the risk model associated to the duration model (ωD)