International Choice Modeling Conference 2011

Modeling the investors' behavior

Thomas Robin **Michel Bierlaire**

2011-07-04

Motivation

- Model how investors are making decisions on stock markets

① Understand the decision process

② Simulate decisions

③ Predict the stock price evolution

Objectives

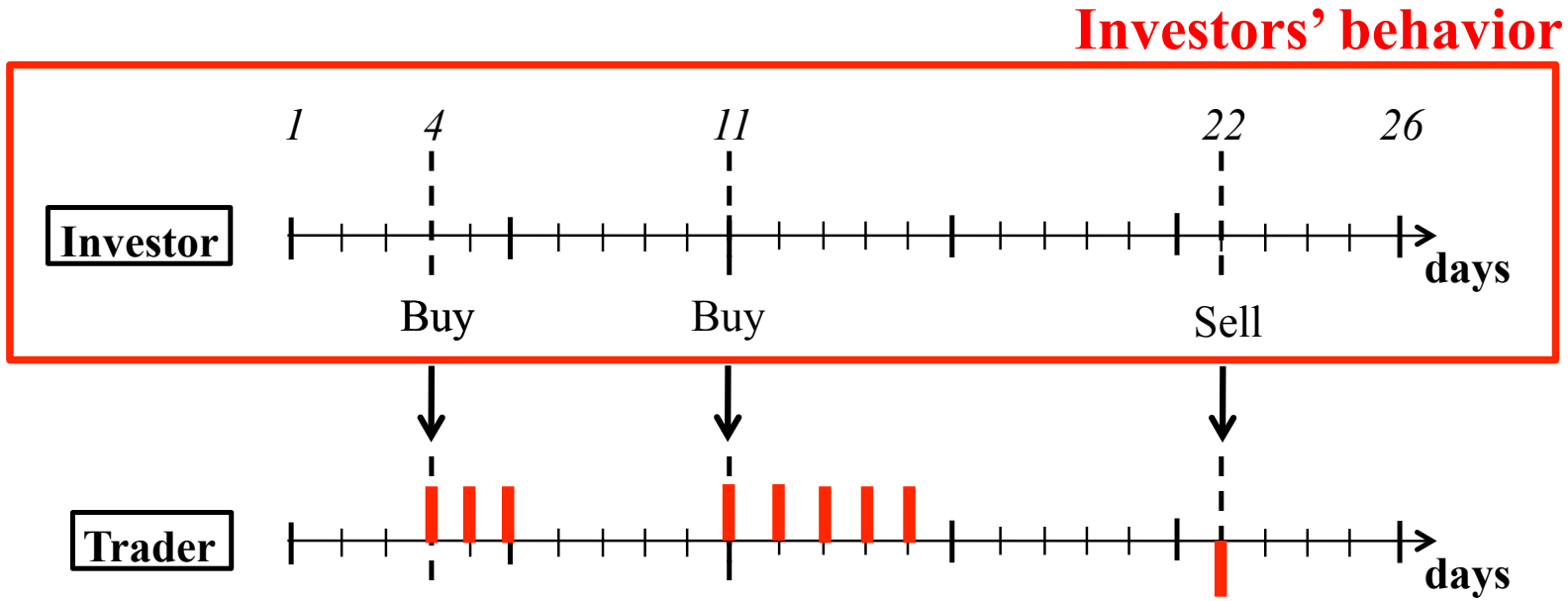
- Model how investors choose an action (buy or sell)
- Model the duration between two consecutive actions

Outline

- Process
- Data
- Correlation analysis
- Model
- Estimation results
- Validation
- Simulation
- Conclusions and Perspectives

Process

- **Investors** initiate **actions** (buy or sell) on **stocks**
- **Traders** translate actions into **transactions**
- Example with 1 company:



Data: raw observations

LOMBARD ODIER
LOMBARD ODIER DARIER HENTSCH
GÉRANTS D'ACTIFS DEPUIS 1796

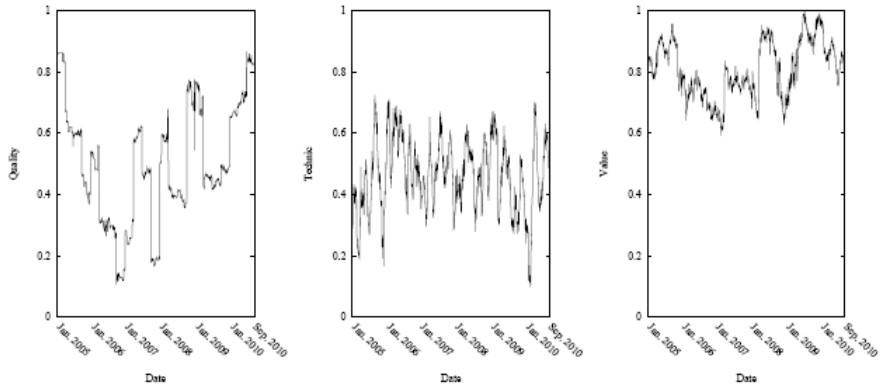
- 25 989 observations of transactions **passed by traders**, on 6 funds (January 2005 – September 2010):

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

- Raw explanatory variables:
  - 5 indicators: **quality** / **sentiment** / **technic** / **value** / **price** (calculated by the bank)
  - 1 market index: **VIX** (interpreted as the volatility of the S&P 500)

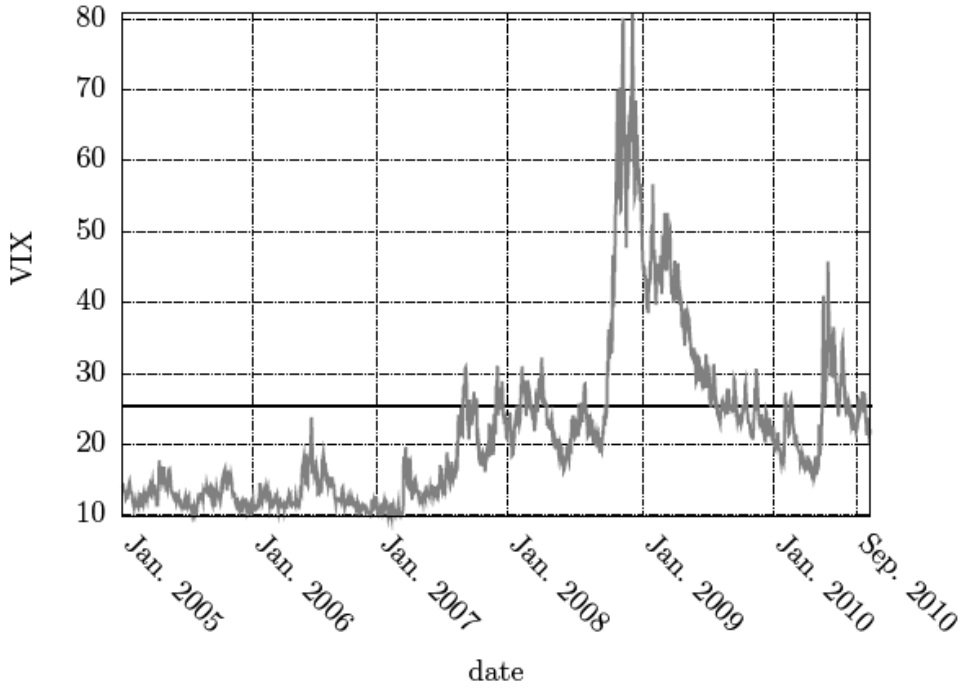
EPFL
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Data: raw explanatory variables



(a) Quality (b) Technic (c) Value

(d) Price (e) Sentiment

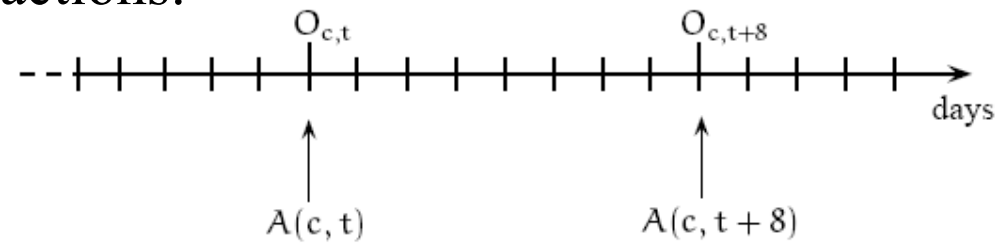


TRANSP-OR

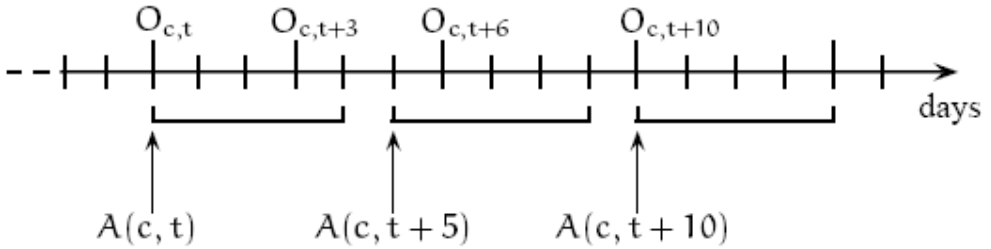
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Data: aggregation

- Aggregation of the transactions:
  - isolated transactions:



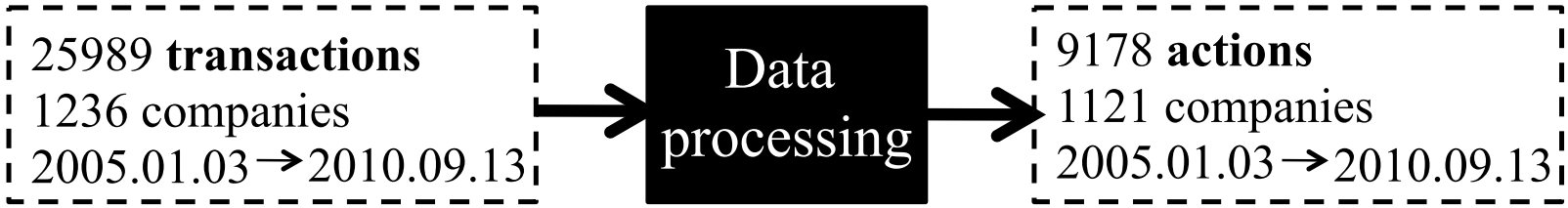
  - neighboring observations:



- The aggregation starts the first « **open**» day of January 2005
- Discarding of the adjustments (actions with a small money amount 25%)

Data: aggregated data

- Observations:



- Actions (buy/sell) passed by traders:

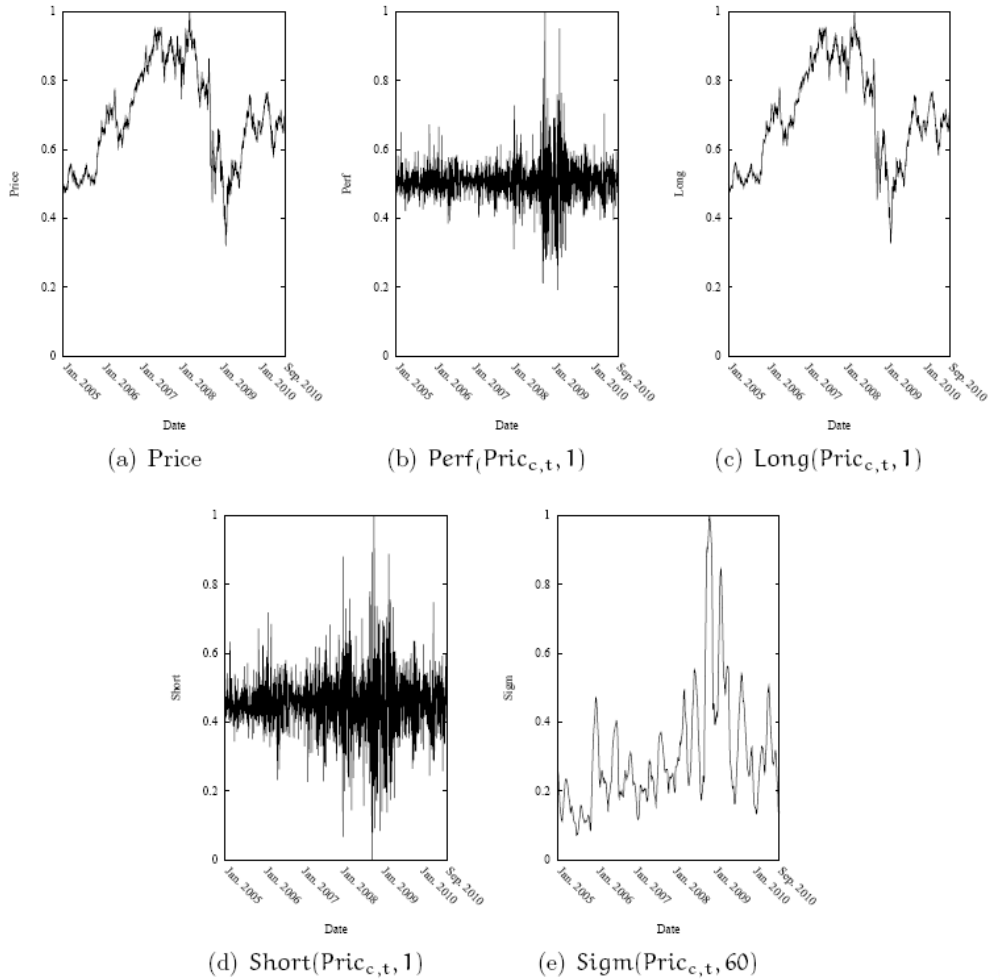
Fund	Nb of decisions	Nb of companies	% buy	% sell
1	1461	145	54.96	45.04
2	913	58	53.34	46.66
3	508	50	59.45	40.55
4	3738	505	51.66	48.34
5	1659	316	43.40	56.60
6	899	175	45.05	54.95

Data: supplementary variables

- Short-term variation: $\mathrm{Short}(x_{c,t}(k), t_H) = x_{c,t}(k) - \mathrm{Long}(x_{c,t}(k), t_H)$
- Long-term variation: $\mathrm{Long}(x_{c,t}(k), t_H) = \frac{1}{t_H} \sum_{l=t-t_H}^{t} x_{c,l}(k)$
- Performance: $\mathrm{Perf}(x_{c,t}(k), t_H) = \frac{x_{c,t}(k) - x_{c,t-t_H}(k)}{x_{c,t-t_H}(k)}$
- Standard-error: $\mathrm{Sigm}(x_{c,t}(k), t_H) = \sqrt{\frac{1}{t_H} \sum_{t=t-t_H}^{t} (x_{c,t}(k) - \mathrm{Long}(x_{c,t}(k), t_H))^2}$

Data: supplementary variables

- Examples:



(a) Price (b) $\text{Perf}_{(}\text{Pric}_{c,t}, 1)$ (c) $\text{Long}(\text{Pric}_{c,t}, 1)$

(d) $\text{Short}(\text{Pric}_{c,t}, 1)$ (e) $\text{Sigm}(\text{Pric}_{c,t}, 60)$

Correlation analysis

- Correlations between the **explanatory variables** and the **actions**
- Codes: Buy = 0 / Sell = 1
- Complete data:

Transform	Variable	t_H	1	2	3	4	5	6
Perf()	Price	1	0.065	0.139	0.127	−0.279	−0.238	−0.329
Short()	Value	3	−0.090	−0.084	−0.099	0.227	0.170	0.277
Sigm()	VIX	360	0.126	0.303	0.449	−0.062	−0.114	0.213
Short()	Technic	1	−0.065	−0.110	−0.113	0.213	0.181	0.257

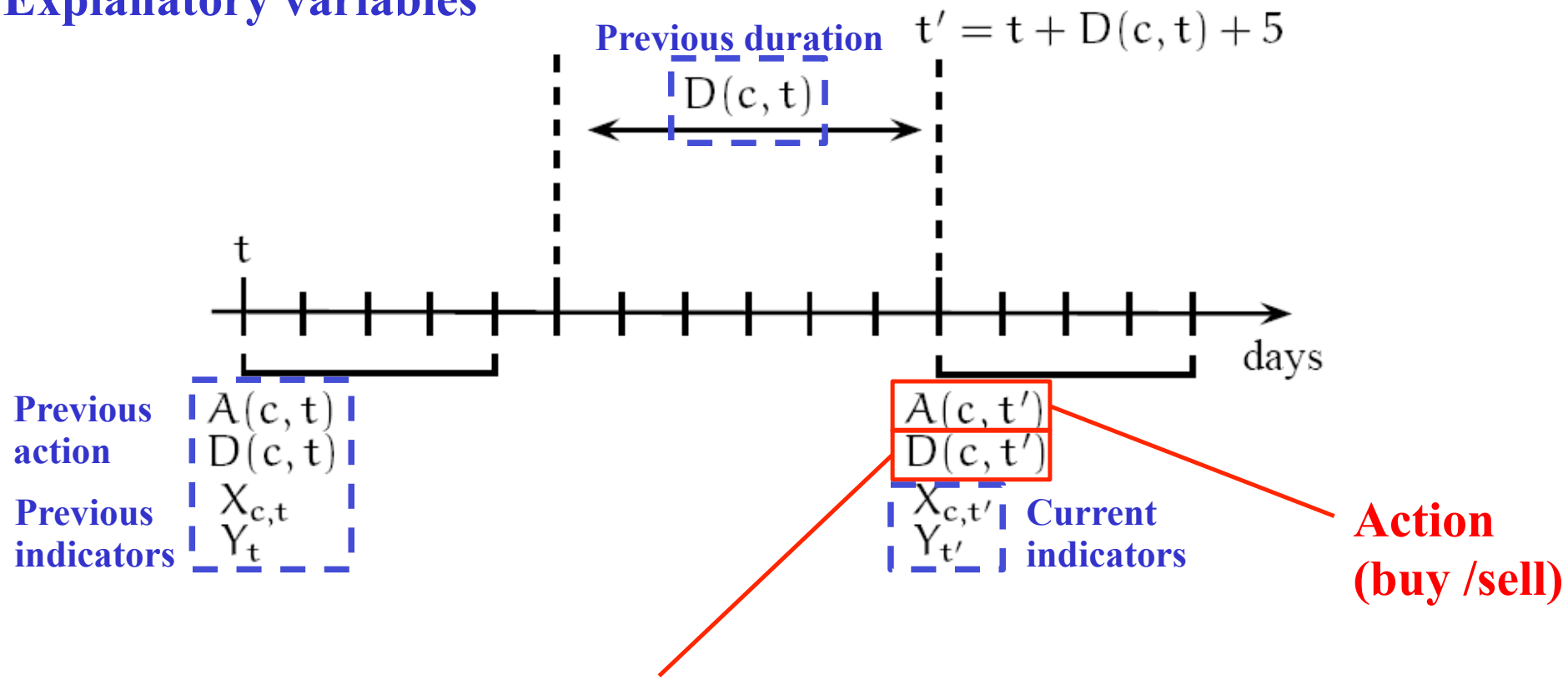
- High VIX:

Transform	Variable	t_H	1	2	3	4	5	6
Short()	Value	60	−0.210	−0.125	−0.200	0.206	0.095	0.232

- 2 fund groups
- Specific information used in risky situations

Model: overview

Explanatory variables



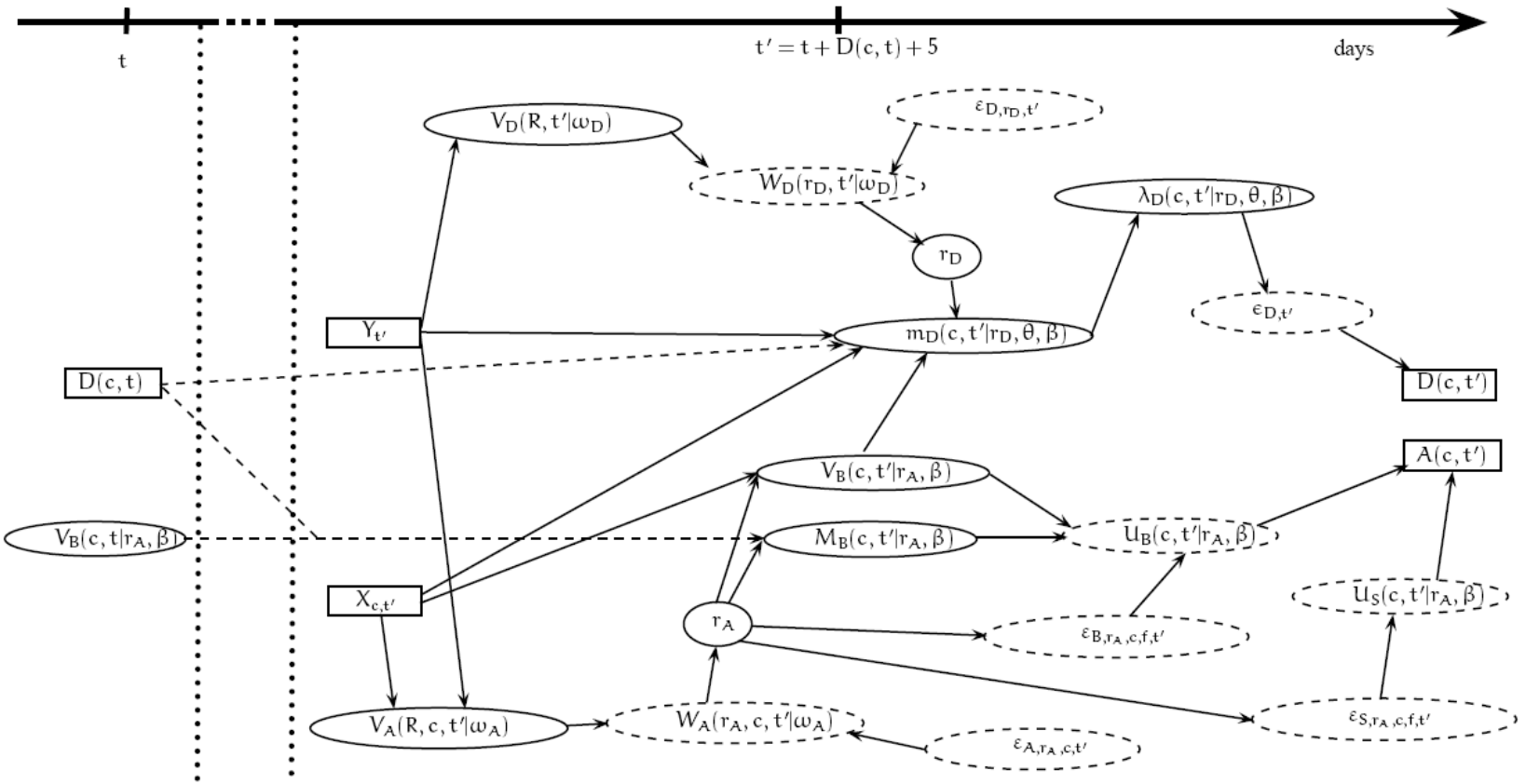
Model: assumptions

- One action per week (5 days) and stock, at most
- The survival time of the action is decided the action day and not revised
- Dynamics:
  - the current **action** influenced by the previous **action**
  - the current **duration** influenced by the previous **action**
- The current duration is influenced by the current action
- The risk perception influences the decisions
- Different perceptions between 2 fund groups

Model: methodologies

- **Action model: discrete choice model**
  - Binary logit model with latent classes
  - Choice set: Buy / Sell
  - Latent classes: risky situation / unrisky situation
  - Dynamic formulation
- **Duration model: survival model**
  - Weibull regression with specification of the scale parameter
  - It accounts for the risk perception
  - Dynamic formulation
  - It accounts for the action model

Model: map of causalities



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Model: action model

- Risk perception model:
  - 2 alternatives for the situations: **risky** and **unrisky**
  - Deterministic measures of the risk:

$$\begin{aligned} V_A(R,c,t'|\omega_A) &= ASC_{W_A} + \omega_{A,1} VIX_{t'} I_{c,g=1} + \omega_{A,2} VIX_{t'} I_{c,g=2} \\ &+ \omega_{A,3} Sigm(Sent_{c,t'}, 5), \end{aligned}$$

  - Random measures:

$$\begin{aligned} W_A(N,c,t'|\omega_A) &= \varepsilon_{A,N,c,t'}, \\ W_A(R,c,t'|\omega_A) &= V_A(R,c,t'|\omega_A) + \varepsilon_{A,R,c,t'}, \end{aligned}$$

  - Random terms:

$$\varepsilon_{A,r_A,c,t'} \overset{i.i.d}{\sim} EV(0,1)$$

  - Probabilities:

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

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

Model: action model

- Choice model: binary logit model with latent classes
  - 2 alternatives: **buy** and **sell**
  - Utilities:

$$\begin{aligned} U_B(c,t'|r_A,\beta) &= V_B(c,t'|r_A,\beta) \\ &+ M_B(c,t'|r_A,\beta) \\ &+ \varepsilon_{B,r_A,c,f,t'}, \\ U_S(c,t'|r_A,\beta) &= \varepsilon_{S,r_A,c,f,t'}, \end{aligned}$$

  - Deterministic utilities:

$$V_B(c,t'|r_A,\beta) = ASC_{B,r_A} + \sum_{g\in\{1,2\}} I_{g,c} \sum_{k=1}^{K_B} \beta_{B,k} \sum_{l=1}^{K_X} I_{B,g,k,l,r_A} X_{c,t'}(l)$$

$$M_B(c,t'|r_A,\beta) = \alpha_{B,r_A} V_B(c,t|r_A,\beta) e^{\lambda_{B,r_A} D(c,t)}$$

  - Random terms: $\varepsilon_{B,r_A,c,f,t'}, \varepsilon_{S,r_A,c,f,t'} \overset{i.i.d}{\sim} EV(0,\mu_f)$
  - Conditional probabilities:

$$P_B(c,t'|r_A,\beta) = \frac{1}{1+e^{-\mu_f V'_B}},$$

$$P_S(c,t'|r_A,\beta) = 1 - P_B(c,t'|r_A,\beta),$$

  - Probabilities:

$$\begin{aligned} P_B(c,t'|\beta,\omega_A) &= P_B(c,t'|N,\beta)P_A(N,c,t'|\omega_A) \\ &+ P_B(c,t'|R,\beta)P_A(R,c,t'|\omega_A), \\ P_S(c,t'|\beta,\omega_A) &= 1 - P_B(c,t'|\beta,\omega_A), \end{aligned}$$

TRANSP-OR

EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Model: duration model

- Risk perception model:
  - 2 alternatives for the situations: **risky** and **unrisky**
  - Deterministic measure of the risk:

$$\begin{aligned} V_D(R, t'|\omega_D) &= ASC_{W_D,1} I_{c,g=1} + ASC_{W_D,2} I_{c,g=2} \\ &+ \omega_{D,1} VIX_{t'} I_{c,g=1} + \omega_{D,2} VIX_{t'} I_{c,g=2}, \end{aligned}$$

  - Random measures:

$$\begin{aligned} W_D(N, t'|\omega_D) &= \varepsilon_{D,N,t'}, \\ W_D(R, t'|\omega_D) &= V_D(R, t'|\omega_D) + \varepsilon_{D,R,t'}, \end{aligned}$$

  - Random terms:

$$\varepsilon_{D,r_D,t'} \overset{i.i.d}{\sim} EV(0,1)$$

  - Probabilities:

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

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

Model: duration model

- Survival model: Weibull regression
  - Dependent variable: **duration between 2 consecutive actions:**

$$D(c,t') = \bar{\epsilon}_{D,t'} \text{ with } \epsilon_{D,t'} \sim W(\lambda_D(c,t'|\theta,\omega_D,\beta),\eta_D)$$

  - Scale of the duration model:

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

with

$$\begin{aligned} m_D(c,t'|\theta,\omega_D,\beta) &= m_D(c,t'|N,\theta,\beta)P_D(N,t'|\omega_D) \\ &+ m_D(c,t'|R,\theta,\beta)P_D(N,t'|\omega_D), \end{aligned}$$

and

$$\begin{aligned} m_D(c,t'|r_D,\theta,\beta) &= \sum_{g\in\{1,2\}} I_{g,c} ASC_{D,r_D,g} + \sum_{g\in\{1,2\}} I_{g,c} \sum_{k=1}^{K_D-1} \theta_{D,k} \sum_{l=1}^{K_X} I_{D,g,k,l,r_D} X_{c,t'}(l) \\ &+ \theta_{D,K_D} Sigm_{VIX_{t'},360} I_{r=N} \\ &+ \alpha_{D,N,1} I_{c,g=1} I_{r_D=N} D(c,t) \\ &+ \alpha_{D,N,2} I_{c,g=2} I_{r_D=N} D(c,t) \\ &+ \alpha_{D,R,2} I_{c,g=1} I_{r_D=R} D(c,t) \\ &+ \theta_{B,r_D} V_B(c,t'|r_D,\beta) \end{aligned}$$

Estimation results: general

- Maximum likelihood estimation using **Python Biogeme**
- Results:

	Action choice model	Duration model	Combined model
Nb parameters	29	32	61
Nb observations	9178	8057	9178
Null Log-likelihood	-6361.705		
Final Log-likelihood	-5635.580	-19711.528	-25347.109
$\bar{\rho}^2/R^2$	0.109	0.048	

- Poor fit of the duration model
- Significant parameters
- Interpretable parameters

Estimation results: action model

- Risk model:

Parameter	ν	Transform	t_H (day)	g	Value	t-test
ASC_{W_A}	1			1,2	-25.327	
$\omega_{A,1}$	VIX			1	1.37	49.41
$\omega_{A,2}$	VIX			2	1.08	31.39
$\omega_{A,3}$	Sentiment	Sigm()	5	1,2	9.29	5.17

- Conditional choice models:

Parameter	ν	Transform	t_H (day)	g	r	Value	t-test
$ASC_{B,N}$	1			1,2	N	-1.39	-4.41
$ASC_{B,R}$	1			1,2	R	-2.01	-4.44
$\beta_{B,1}$	Price	Long()	1	1	N	-2.31	-3.15
$\beta_{B,2}$	Price	Perf()	1	2	N,R	6.18	4.19
$\beta_{B,3}$	Price	Short()	1	2	N	-1.02	-3.15
$\beta_{B,4}$	Price	Short()	1	1,2	R	0.916	1.78
$\beta_{B,5}$	Quality	Long()	1	1,2	N	-0.428	-2.57
$\beta_{B,6}$	Quality	Short()	1	1,2	N	0.723	2.03
$\beta_{B,7}$	Sentiment	Long()	1	2	N	0.659	2.78
$\beta_{B,8}$	Technic	Long()	1	1	R	1.06	1.84
$\beta_{B,9}$	Technic	Long()	1	1	N	1.02	2.51
$\beta_{B,10}$	Technic	Long()	1	2	R	-0.805	-2.42
$\beta_{B,11}$	Technic	Short()	1	1	R	1.80	2.45
$\beta_{B,12}$	Technic	Short()	1	1	N	2.78	4.28
$\beta_{B,13}$	Technic	Short()	1	2	R	1.30	3.40
$\beta_{B,14}$	Technic	Short()	1	2	N	-1.47	-2.00
$\beta_{B,15}$	Value	Short()	1	1,2	N	0.625	2.43
α_N				1,2	N	-0.222	-5.43
α_R				1,2	R	-0.312	-6.77
$\lambda_{B,N}$				1,2	N	-0.0156	3.79
$\lambda_{B,R}$				1,2	R	-0.00946	5.27
μ_1				1,2	N,R	1	
μ_2				1,2	N,R	1.23	5.07
μ_3				1,2	N,R	1.58	4.21
μ_4				1,2	N,R	2.93	3.16
μ_5				1,2	N,R	1.43	2.77
μ_6				1,2	N,R	2.46	3.48

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Estimation results: duration model

- Risk model:

Parameter	ν	Transform	t_H (day)	g	Value	t-test
$ASC_{W_D,1}$	1			1	-7.48	-1.59
$ASC_{W_D,2}$	1			2	-4.93	-3.76
$\omega_{D,1}$	VIX			1	0.377	2.10
$\omega_{D,2}$	VIX			2	0.263	4.53

- Conditional duration models:

Parameter	ν	Transform	t_H (day)	g	r	Value	t-test
$ASC_{D,N,1}$	1			1	N	3.54	5.49
$ASC_{D,N,2}$	1			2	N	3.36	10.71
$ASC_{D,R,1}$	1			1	N	0.242	1.54
$ASC_{D,R,2}$	1			2	N	0.398	2.93
$\theta_{D,1}$	Price	Short()	60	1	N	-1.22	-3.47
$\theta_{D,2}$	Price	Sigm()	60	1	N	1.52	1.78
$\theta_{D,3}$	Quality	Long()	60	1	N	-1.08	-3.03
$\theta_{D,4}$	Quality	Short()	60	1	R	0.960	1.80
$\theta_{D,5}$	Quality	Short()	60	2	R	-0.661	-1.54
$\theta_{D,6}$	Quality	Short()	60	1	N	-1.43	-1.96
$\theta_{D,7}$	Sentiment	Long()	60	2	R	-0.716	-3.29
$\theta_{D,8}$	Sentiment	Short()	60	2	R	0.990	3.72
$\theta_{D,9}$	Technic	Long()	60	2	R	1.42	5.38
$\theta_{D,10}$	Technic	Long()	60	1	N	-1.18	-2.36
$\theta_{D,11}$	Technic	Short()	60	1	R	1.79	3.19
$\theta_{D,12}$	Technic	Sigm()	60	2	N	-1.48	-3.91
$\theta_{D,13}$	Value	Short()	60	2	R	1.90	4.84
$\theta_{D,14}$	Value	Short()	60	1	N	2.25	2.60
$\theta_{D,15}$	Value	Sigm()	60	2	N	-0.613	-1.72
$\theta_{D,16}$	VIX	Sigm()	360	2	N	-2.05	-4.99
$\theta_{B,N}$				1,2	N	-0.350	-2.27
$\theta_{B,R}$				1,2	R	-0.261	-2.53
$\alpha_{D,N,1}$				1	N	7.09	4.77
$\alpha_{D,R,1}$				1	R	5.29	3.25
$\alpha_{D,N,2}$				2	R	2.23	1.84
η_D						0.530	82.26

Estimation: interpretation of the parameters

- **Overall**
  - Different perceptions between the 2 fund groups
  - Different perceptions under the 2 risk situations
  - The risk is characterized by the VIX and the sentiment variations
  - Perception of short and long-term variations of quality, value, technic and price
- **Action model**
  - Explanatory variables calculated with a time horizon of 1 day
  - Investors have not tendency to perform consecutively similar actions
- **Duration model**
  - Lower duration in risky situations
  - Explanatory variables calculated with a time horizon of 60 days
  - The current duration tends to increase when the previous duration increases
  - When investors buy stocks, they do not wait long for performing a new action

Validation

- Evaluation of the prediction capabilities of the model
- Cross-validation:
  - Division of the data in 5 subsets
  - Estimation on 4 subsets /Simulation on 1 subset

Periods:

Validation set	1	2	3	4	5
Starting date	2005.01.03	2006.03.02	2007.04.25	2008.06.05	2009.07.24
Ending date	2006.03.01	2007.04.24	2008.06.04	2009.07.23	2010.09.13
Nb actions	1363	1511	1766	2149	2399

Results:

Experience	1	2	3	4	5	
R_e^2	0.034	0.047	-0.003	-0.003	-1.790	Duration model
Estimated R^2	0.053	0.065	0.010	0.025	-0.464	
Predicted $\mathcal{L}$	-394.436	-431.504	-1095.921	-1269.583	-1588.811	Action model
Estimated $\mathcal{L}$	-383.671	-420.055	-1052.993	-1220.002	-1353.67	

→ **- The duration model is not stable**
- Generality of the action model

Simulation: prediction of the future actions

- Evaluation of the **action model** as decision-aid tool
- Work on the last experience of the cross-validation
  - **Estimation** on the period 2005.01.03 ⟶ 2009.07.23
  - **Prediction** on the period 2009.07.24 ⟶ 2010.09.13
- Simulation:

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

Simulation: results

- Calculate relative errors: $$Err = \frac{nb\_buy\_obs - n\_buy\_sim}{n\_buy\_obs + n\_sell\_obs} \times 100$$

- Results:

Month	Simul. 1	Simul. 2	Simul. 3	Simul. 4	Simul. 5
Jul. 09	1.92	-7.69	-9.62	-1.92	5.77
Aug. 09	5.94	12.38	12.38	13.37	6.44
Sep. 09	8.85	1.04	-4.17	3.65	0.00
Oct. 09	-4.38	-7.97	-2.39	-2.39	-0.80
Nov. 09	-1.65	-2.88	-4.94	-3.29	0.41
Dec. 09	0.59	-5.92	-0.59	0.59	10.65
Jan. 10	-1.32	-0.66	1.32	0.00	-0.66
Feb. 10	0.68	5.48	8.90	0.68	1.37
Mar. 10	7.78	6.11	5.00	3.89	5.56
Apr. 10	25.74	17.82	21.78	23.27	19.80
May 10	5.33	0.00	5.92	1.18	-0.59
Jun. 10	7.87	3.37	0.00	-8.99	1.12
Jul. 10	-8.05	-16.11	-16.78	-12.75	-18.79
Aug. 10	-0.58	-4.07	8.72	-4.07	2.91
Sep. 10	-18.75	9.38	3.13	-9.38	6.25

Highest value = 25.74% / highly satisfactory results

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Conclusions and Perspectives

- Conclusions:
  - Correlation analysis
  - Specification and estimation of the model on real data
  - Validation of the prediction capabilities of the model
  - Simulation with the action model
- Perspectives:
  - Improve the risk perception (more than 2 classes)
  - Improve the duration model
  - Predict the evolution of the stock market

Thanks