ST4 Enseignement d'intégration
Calibration de modèles d'agents

damien.challet@centralesupelec.fr

26th May 2023

This week aims

1. Calibration of ABMs to financial data
2. Prediction of financial prices with ABMs

This week programme

1. Friday+Tuesday: calibration of a simple model
2. Wednesday: empirical trader behavior
3. Thursday: prediction
4. Friday:
   4.1 morning: oral presentation preparation
   4.2 afternoon: oral presentation.

- Mark on pdf of oral presentation + presentation

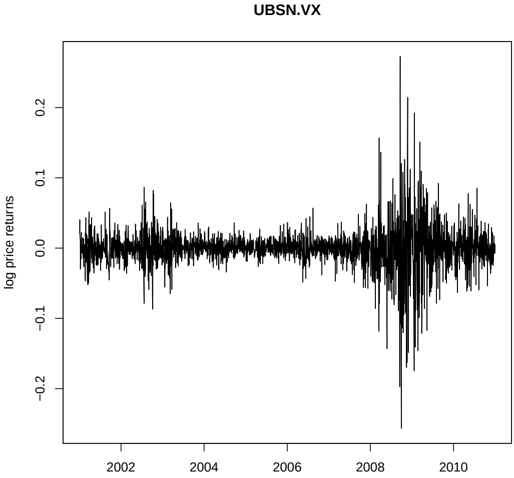
ABMs: what for?

1. Reproduce empirical facts
   $\rightarrow$ Find fundamental mechanisms — science
2. What if — policies
3. Prediction — planning, money

- If 1. works, calibration makes sense
- 2. and 3. require calibration

Empirical facts: price returns

clustered volatility | heavy-tailed price returns



ABM calibration levels

Parameters of the model from

1. Global quantities
   - stationary distributions
   - dynamical quantities: correlations, etc
2. Internal state of the model
   - find the most likely state at each time step
3. Individual behaviour
   - inventory evolution: which strategy?
   - inactivity?

ABM: framework

- $\xi \in \mathbb{R}^K$: K parameters
- Output: r_t e.g. price returns
- Input: similar time series, either from an ABM or from real system

3 main questions:

1. If r_t from ABM, can one recover ξ?
2. If r_t from real system, ξ?
3. What does one learn about reality from fitted ABM?

Level 1: best case

Hypothesis:

1. Exact results for the ABM
2. Time series $\{r_t\}_{1 \leq t \leq T}$ comes from the ABM
3. r_t i.i.d.
4. Stationary distribution $P(r;\xi)$ is known and has $J = K$ parameters

- Maximum Likelihood Estimation (MLE)

$$\begin{aligned}\hat{\xi} &= \arg\max_{\xi} \prod_{t=1}^{T} P(r_t;\xi) \\ &= \arg\max_{\xi} \sum_{t=1}^{T} \log P(r_t;\xi)\end{aligned}$$

Level 1: second best case

- Stationary distribution $P(r;\xi)$ is known and has $J < K$ parameters
- MLE $\rightarrow$ J parameters
- $K - J$ parameters: from dynamical quantities
  - auto-correlation of r_t (e.g. at lags 1, 5, 19)
  - auto-correlation of r_t^2(e.g. at lags 2,8, 32)
  - etc.

Level 1: moment matching

If the shape distribution of empirical data $P(r;\theta)$ is known, $\theta \in \mathbb{R}^K$, write

$$\begin{aligned} E(r) &= f_1(\theta) \\ E(r^2) &= f_2(\theta) \\ &\vdots \\ E(r^K) &= f_K(\theta) \end{aligned}$$

and solve these equations for θ from the moment estimates

$$\hat{\mu}_k,\ k = 1, \cdots, K$$

Level 1: Generalised moments method (GMM)

- ABM output: $r_{ABM}(t)$
- Idea: find parameters $\hat{\xi}$ such that

$$E(r^k_{ABM};\xi) \simeq E(r^k), \quad k \in K$$

- Define distance D_W

$$D_W(\xi) = \sum_{k,j=1}^{K} [E(r^k_{ABM};\xi) - E(r^k)]\, W_{kj} [E(r^j_{ABM};\xi) - E(r^j)]$$

$W : \ I$ or inverse of moments covariance.

and find

$$\hat{\xi}_{ABM} = \arg\min_{\theta} D(\xi)$$

Optimal GMM

Hanson (1982) [paper]

Theorem

If $W_{kj} = V_{kj}^{-1*}$ where V^* is the moments covariance matrix for the optimal parameters, then the GMM estimator is optimal (minimum variance)

In practice, recommended iterative procedure:

1. start with $W^{(0)}$
2. find optimal parameters $\xi^{(0)}$
3. update $W^{(1)}$ from $(W^{(0)}, \xi^{(0)})$
   - simulate many times model with $\xi = \xi^{(0)}$
   - compute moment covariance between the runs
4. repeat 2.&3. until $|\xi^{(n)} - \xi^{(n-1)}| < \epsilon$

Level 2: stationary and dynamical quantities

Dynamics of ABMs: more than moments of r

- Tail exponent of distributions (returns, volatility)
- Autocorrelation at lag 1
- Full shape of autocorrelation function
- (moments)

Level 2: GMM

- Let c_k be a quantity of interest
- Measure it in real data $\rightarrow \hat{c}_{k,r}$
- Measuring it in ABM(ξ) $\rightarrow \hat{c}_{k,ABM}(\xi)$
- Define distance

$$D_W(\xi) = \sum_{i,k=1}^{K} (\hat{c}_{i,ABM} - \hat{c}_{i,r}) W_{ik} (\hat{c}_{k,ABM} - \hat{c}_{k,r})$$

and find $\hat{\xi}$ which minimises W

Example: Franke-Westerhoff [link]

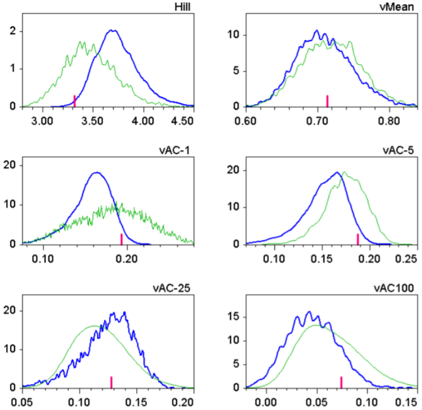


Figure 3: Distributions of selected moments (based on empirical data).

Note: The bold (blue) line represents the distributions of the moments resulting from the MC simulations of θ^o, the solid (green) line depicts the corresponding bootstrap distributions from $\{r_t^{emp}\}$; both with a time horizon of 6750 days. The (red) vertical bar indicates the empirical moments.

How hard is parameter estimation?

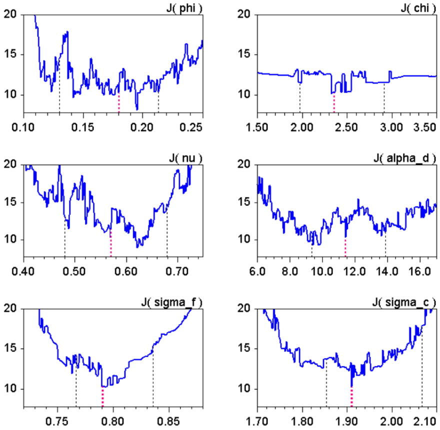


Figure 5: Graphs of the functions $\theta_i \mapsto J^o(\theta_i)$, for $\theta_i = \phi, \chi, \nu, \alpha_d, \sigma_f, \sigma_c$.

Note: Middle (red) dashed line depicts the benchmark value θ_i^o of the parameter, the two outer dashed lines indicate the 2.5% and 97.5% quantiles of the estimations presented below. The underlying time horizon is $S = 68660$.

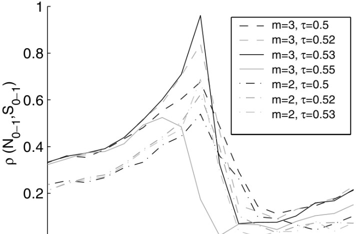
Level 3: price dynamics matching

- Given $r(t)$, find ξ such that distance between $r(t)$ and $r_{ABM}(t)$ is minimal
- Distances
  1. MSE
  $$\frac{1}{T}\sum_t [r_{ABM}(t) - r(t)]^2$$
  2. 1−Correlation (e.g. Kendall)
  $$1 - C(r_{ABM}, r)$$
- Equivalent to reverse engineering

Fitting r_t: reverse engineering a fully deterministic model

Lamper, Howison, Johnson (2002): fit the Minority game $(N \cdot S \cdot 2^M)$ parameters

1. Generate a reference model: BLACK BOX
2. Find parameters set:
   2.1 sweep over all of them.
   For a given set of parameters
      2.1.1 generate many games
      2.1.2 compute cross-correlation$\sim E(A_{\text{guess}} A_{\text{black box}})$
      2.1.3 keep the game with highest cross-correlation



Level 3: Bayesian estimation of ABMs

- From ABM dynamics and parameters $\xi \to r_{ABM,\xi}$
- This yields
$$P(r_{ABM}|\xi)$$
- Calibration means
$$\hat{\xi} = \arg\max_{\xi} P(\xi|r_{ABM})$$
- Bayes theorem
$$P(\xi|r_{ABM}) = \frac{P(r_{ABM}|\xi)}{P(\xi)}$$
- State-of-the-art: at each time step t, update prior on ξ from r_t

Level 3: Bayesian estimation of ABMs

Bayesian estimate of FW parameters, Bertschinger et al. (2018)

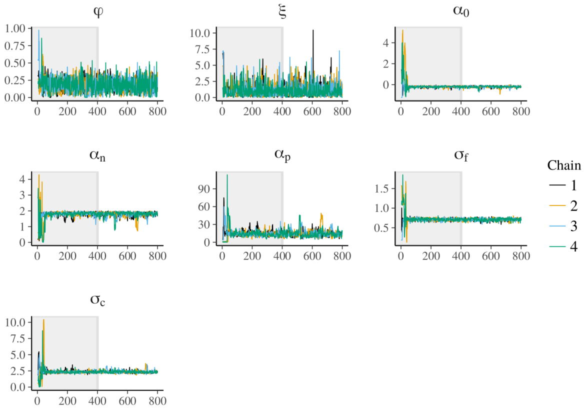


Figure 3: Trace plot for model parameters $\phi, \xi, \alpha_0, \alpha_n, \alpha_p, \sigma_f$ and σ_c. Note that all chains appear to have converged to the same posterior distribution after just about 50 samples.

Bayesian calibration

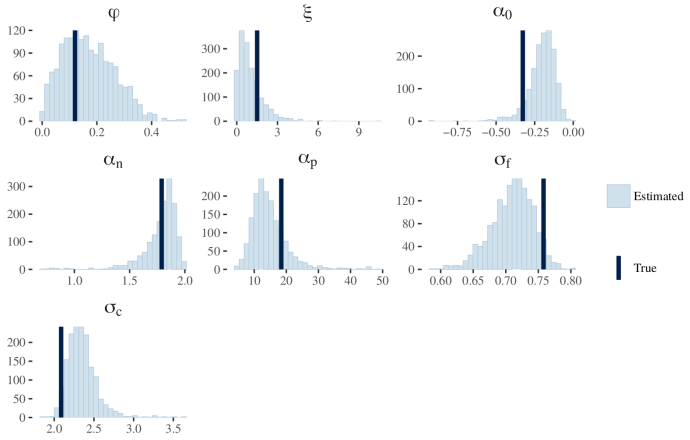


Figure 4: Plot of posterior densities for parameters $\phi, \xi, \alpha_0, \alpha_n, \alpha_p, \sigma_f$ and σ_c. The true values are well covered by the posterior distributions.

Calibration Level 4: fit agent behaviour

From the actions of individual real traders

1. Macro approach
   - A_t: excess demand of investors between t and $t+1$
   - Find f:
     $$r_{t+1} = f(A_t) + noise$$
     and match with model dynamics
2. Micro approach
   - $a_{i,t}$: excess demand of agent i between t and $t+1$
   - Find how $a_{i,t}$ depends on the past
   - Equivalent to finding strategy of each agent
   - or to find which strategy is the most likely

Calibration Level 4: example

Lillo et al. (2007) [link]

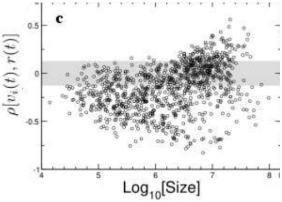
- Inventory (number of shares) of trader i: $V_i(t)$
- Inventory variation

$$v_i(t) = V_i(t) - V_i(t-1)$$

- Compute correlation

$$\rho[v_i(t), r(t)]$$

and classify traders



	2001		2002		2003		2004	
Reversing	43	(52%)	39	(49%)	42	(52%)	37	(51%)
Uncategorized	28	(34%)	31	(39%)	31	(38%)	29	(40%)
Trending	11	(13%)	10	(12%)	8	(9.9%)	6	(8.3%)
Total	82		80		81		72	

II. Prediction

Rationale

- A good ABM captures fundamental elements of market dynamics.
- Its internal state at time t contains some information about the future.

Agenda

1. Fit the model to data $\rightarrow$ parameters
2. Infer state at time t
3. Run the model a few time-steps ahead
   $\rightarrow$ prediction of $P(r(t+k))$ or $P(|r(t)|)$
4. (profit)

1. Calibration (Friday-Tuesday)

Majewski et al. (2020) [link].

Journal of Economic Dynamics & Control 112 (2020) 103791

Contents lists available at ScienceDirect

Journal of Economic Dynamics & Control

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

ELSEVIER

Co-existence of trend and value in financial markets: Estimating an extended Chiarella model☆

Adam A. Majewski*, Stefano Ciliberti, Jean-Philippe Bouchaud

Capital Fund Management 23 rue de l'Université Paris 75007, France

1. Calibration

Majewski et al. (2020) [link].

Consider the discrete version of the model (Eqs. (2.3)): log price p_t

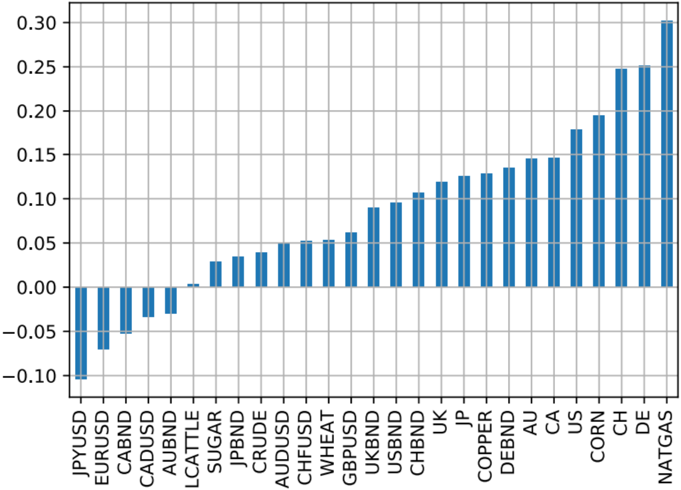
$$\begin{aligned} p_{t+1} - p_t &= \kappa_1(v_t - p_t) + \beta \tanh(\gamma m_t) + \epsilon_{t+1} \\ m_t &= (1-\alpha)m_{t-1} + \alpha(p_t - p_{t-1}) \\ v_t &= v_{t-1} + g + \eta_t \end{aligned}$$

where

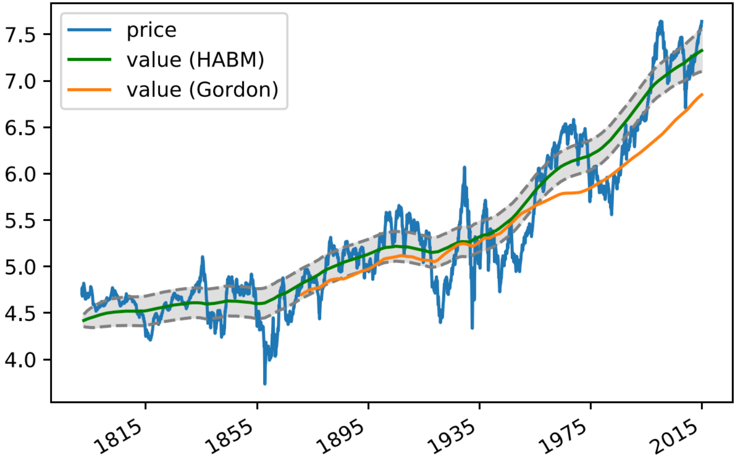
$$\begin{aligned} m_t &: \text{ exponential moving average of the trend} \\ v_t &: \text{ fundamental value} \\ \eta_t &\sim \mathcal{N}(0, \sigma_V^2) \\ \epsilon_t &\sim \mathcal{N}(0, \sigma_N^2) \end{aligned}$$

1. Calibration: model properties

correlation between predicted returns and future returns



1. Calibration: model properties



1. Calibration

1. Replace the dynamics of the fundamental log-value by an exponentially moving average of past prices

$$v_t = (1 - \lambda) v_{t-1} + \lambda p_t$$

2. Implement a GMM with as many quantities as parameters. Choose the quantities (moments, dynamic quantities, tail exponents, statistical tests on price return Gaussianity, etc.) that are fast to compute

3. Choose a single set of parameters and perform the GMM 100 times.
4. Plot the densities of the each parameter. Comment on their sloppiness.
5. What is the standard deviation of each parameter? What is the RMSE? Comment on the precision of each parameter (see also 4.)

Bonus topic

1. Add a non-linear term to the fundamentalists demand: the model becomes

$$\begin{aligned} p_{t+1} - p_t &= \kappa_1(v_t - p_t) + \kappa_3(v_t - p_t)^3 + \beta \tanh(\gamma m_t) + \epsilon_{t+1} \\ m_t &= (1-\alpha)m_{t-1} + \alpha(p_t - p_{t-1}) \\ v_t &= (1-\lambda)v_{t-1} + \lambda p_t \end{aligned}$$

2. Modify your GMM algorithm
3. Choose a single set of parameters and perform the GMM 100 times.
4. Plot the densities of the each parameter. Comment on their sloppiness.
5. What is the standard deviation of each parameter. Comment on the precision of each parameter.

Technical hints

- %time in a cell tells you how much time it took to run
- %lprun of a function: line profiling, how much time each line of a function takes to run
- Numpy = C: fast
- Numba may help

2. Investor behaviour: Wednesday

1. Retail investors
   - not great at portfolio diversification
   - most lose money by doing too much
2. Institutional investors
   - not great at having new ideas
   - most make money by not doing much

2. Retail investors

The Behavior of Individual Investors*

Brad M. Barber[a] **and Terrance Odean**[b]
[a]Graduate School of Management, University of California, Davis, Davis, CA 95616, USA. Tel.:+1 (530) 752 0512
[b]Haas School of Business, University of California, Berkeley, Berkeley, CA 94720, USA. Tel.:+1 (510) 642 6767

[link]

2. Retail investors

1. The Performance of Individual Investors
   - **1.1** The Average Individual
     - *1.1.1 Long-Horizon Results*
     - *1.1.2 Short-Horizon Results*
     - *1.1.3 Market vs. Limit Orders*
   - **1.2** Cross-Sectional Variation in Performance
2. Why do Individual Investors Underperform?
   - **2.1** Asymmetric Information
   - **2.2** Overconfidence
   - **2.3** Sensation Seeking
   - **2.4** Familiarity
3. The Disposition Effect: Selling Winners and Holding Losers
   - **3.1** The Evidence
   - **3.2** Why Do Investors Prefer to Sell Winners?
4. Reinforcement Learning
5. Attention: Chasing the Action
6. Failure to Diversify
7. Are Individual Investors Contrarians?

Retail investors: contrarians or trend-followers?

- Definition: trend-following if

$$Corr[(p_t - p_{t-T})v_{i,t}] > 0$$

  where $v_{i,t} = V_{i,t} - V_{i,t-1}$ is the difference of shares in the holdings of agent i.
- Definition: contrarian if

$$Corr[(p_t - p_{t-T})v_{i,t}] < 0$$

- Fact 1: for T small (days), most individual traders are contrarians
  - buy after a price drop
  - sell after a new price high.
- Fact 2: longer-term trend-following expectations

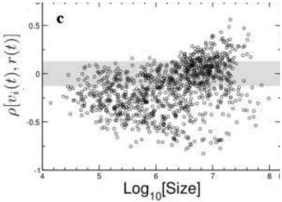
Example

Lillo et al. (2007) [link]

- One-day price return ($T = 1$)
- Compute correlation

$$\rho[v_i(t), r(t)]$$

and classify traders



Expectations from experiments

da Gama Batista *et al.* (2018) [link]: experiments + expectations

- before buying,
$$E(r_{t+1}) < 0$$
- before selling,
$$E(r_{t+1}) > 0$$
- when not holding a position
$$E(r_{t+1}) = 0$$
- when holding a position
$$E(r_{t+1}) = E(r_{market})$$

2. Investor behaviour: data

Guttiérez-Roig *et al.* (2018) [link]

Gutiérrez-Roig et al. *EPJ Data Science* (2019) 8:10
https://doi.org/10.1140/epjds/s13688-019-0188-6
EPJ.org
EPJ Data Science
a SpringerOpen Journal
REGULAR ARTICLE Open Access
Check for updates

Mapping individual behavior in financial markets: synchronization and anticipation

Mario Gutiérrez-Roig[1,4*], Javier Borge-Holthoefer[2], Alex Arenas[3] and Josep Perelló[4,5]

data: [link]

Number of shares in individual investors at the end of each day

2. Investor behaviour: exploration

1. Read the part of the README file that describes the data.
2. Choose an asset, download the investors and price/volume files.
3. Compute the amount of money invested by each investor, denoted by m_i and plot 1-ECDF(m). Is this distribution heavy-tailed? Do you think that it corresponds to what the README file claims?
4. Compute the total number of changes of invested volume (in shares) for each investor and plot its 1-ECDF. What kind of distribution do you find?

2. Investor behaviour: the influence of past price returns

Follow Lillo et al. (2007) [link]

1. Classify the investors according to the correlation between their change of inventory and the same-day return into neutral, trend-following, and mean-reverting.
2. Idem, but cut the data intro two periods of roughly equal length. Compute the transition probability that an investor changes from one class to another. Compute the whole transition table.
3. Idem, but for the correlation between the change of inventory and the price return of the next day. The aim is to see what class of investor is able to really predict future price moves.

2. Investor behaviour: the influence of past price returns

Generate 10 signals of $\kappa(v_t - p_t)$ and 10 of $\beta \tanh(\Gamma m_t)$ (or m_t) as defined in Majewski *et al.* model, for reasonable values of parameters.

1. For each agent, find which signal has the highest correlation with her change of inventory over the whole period. Plot the histogram of the frequency of selection of each signal.
2. Idem, but for two periods. Compute the transition table.

3. Prediction with an ABM

Recipe

1. Choose a model
2. Calibration window $[t-T, t] \rightarrow$ parameters ξ_t
3. Predict r_{t+1} from the state of the model at time t; act accordingly
4. Update the state of the model
5. Repeat 3 and 4
6. If the system is non-stationary, fit the model every δt timesteps (rolling windows).

3. Prediction: ABM $\rightarrow$ ABM, stationary

1. Use Majewski *et al.* type of model (including possibly more signals). Take realistic parameters. Simulate $2T$ timesteps.
2. Calibrate the model with the first T timesteps.
3. How can one infer the state of the model at time $t = T$?
4. For the next T timesteps, find the state at time t, predict the state at time $t + 1$.
5. Check your success rate (precision, hit ratio).
6. (optional) Check if lagging trend followers demand improves the predictions.

3. Prediction: ABM $\to$ real world

1. Take a long-enough real price time-series.
2. Calibrate the model with the first T timesteps.
3. Infer the state of the model at time $t = T$.
4. For the next δt timesteps, find the state at time t, predict the state at time $t + 1$; compute the precision and the hit ratio.
5. Shift the calibration window by δt; redo 2-4; repeat 2-5 until the end of the timeseries.
6. By transforming the prediction into an action, plot the cumulative performance of your prediction. Note: transaction costs $\simeq 0.01\%$ for equities, 0.3% for cryptos.

3. Prediction: some caveats

1. Majewski *et al.* use monthly returns
   1.1 returns are mostly Gaussian
   1.2 value vs trend may be only found that this scale

2. Daily data?

3. Intraday data: very variable activity $\rightarrow$ focus or equal-volume time slots.

4. Computation time: you have 4-6 laptops, so about 16 cpu cores.

4. Presentation

1. In English. Indeed. Avoid writing down full sentences like this one here as it is really hard to listen and read something else at the same time, especially when the sentence is about science.

2. 20 minutes talk, 10 minutes questions

3. 4*5 minutes: 3 project parts + discussion about group and project management

4. Friday 10 June, EB.118, 14:00-17:00
   - group n @ 14:00+$(n-1)\times$30 minutes