ISTITUTO
DI ECONOMIA

Market selection and learning under model misspecification

Giulio Bottazzi[1] **Daniele Giachini**[1] Matteo Ottaviani[2]

[1] Institute of Economics and Department L'EMbeDS,
Scuola Superiore Sant'Anna, Pisa, Italy.

[2] DZHV, Berlin, Germany; Scuola Normale Superiore, Pisa, Italy.

EEA Annual Congress, August 31st, 2023, Barcelona.

Introduction

- **Predictive accuracy**, **market selection**, and **asset pricing** are deeply connected.

Introduction

- **Predictive accuracy**, **market selection**, and **asset pricing** are deeply connected.
- Intuition in pure-exchange complete-markets Arrow-Debreu economies under general equilibrium, bounded endowments, and discount factor homogeneity:

  1. i learns the **true model** (data generating process) better than j,
  2. i allocates more wealth than j to events that actually are *more likely* to happen,
  3. i becomes richer than j,
  4. i drives the price toward the truth.

Introduction

- When the learning problems are **correctly specified**:
  - a Bayesian agent learns the true model,
  - it drives anyone who forecasts differently out of the market and sets prices,
  - thus, long-run prices are consistent with rational expectations,
  - hence, the ecology of traders and selection dynamics do not matter.

Introduction

- When the learning problems are **correctly specified**:
  - a Bayesian agent learns the true model,
  - it drives anyone who forecasts differently out of the market and sets prices,
  - thus, long-run prices are consistent with rational expectations,
  - hence, the ecology of traders and selection dynamics do not matter.

- Correct specification is a strong assumption...
- What happens if correct specification does not hold?

Introduction

- Box (1976) : “*all models are wrong*”,

Introduction

- Box (1976) : “*all models are wrong*”,

- Cox (1995): “*...does not seem helpful just to say that all models are wrong. The very word “model” implies simplification and idealization. The idea that complex physical, biological or sociological systems can be exactly described by a few formulae is patently absurd*”.

Introduction

- Box (1976) : "*all models are wrong*",

- Cox (1995): "*...does not seem helpful just to say that all models are wrong. The very word "model" implies simplification and idealization. The idea that complex physical, biological or sociological systems can be exactly described by a few formulae is patently absurd*".

- All models agents use to learn are approximations: **model misspecification** (Hansen, 2014; Fudenberg et al., 2017; Marinacci and Massari, 2019; Cerreia-Vioglio et al., 2020; Hansen and Sargent, 2022).

Introduction

- Under model misspecification, Bayesian updating loses its formal justification (Massari, 2021).

Introduction

- Under model misspecification, Bayesian updating loses its formal justification (Massari, 2021).
- Evidence that traders showing **cognitive biases** may dominate a Bayesian learner (Massari, 2020; Antico et al., 2023).

Introduction

- Under model misspecification, Bayesian updating loses its formal justification (Massari, 2021).
- Evidence that traders showing **cognitive biases** may dominate a Bayesian learner (Massari, 2020; Antico et al., 2023).
- Consider an ecology of traders characterized by heterogeneous learning rules,
  - how are the emerging selection outcomes influenced by model misspecification?
  - Can particular survival learning mechanisms be identified?
  - How general are they?

Introduction

- Under model misspecification, Bayesian updating loses its formal justification (Massari, 2021).
- Evidence that traders showing **cognitive biases** may dominate a Bayesian learner (Massari, 2020; Antico et al., 2023).
- Consider an ecology of traders characterized by heterogeneous learning rules,
  - how are the emerging selection outcomes influenced by model misspecification?
  - Can particular survival learning mechanisms be identified?
  - How general are they?
- We study selection outcomes considering 4 learning processes (Bayes, underreaction, moving average, limited memory Bayes) and 2 cases of model misspecification (parametric and structural).

The Model

Arrow-Debreu economy in discrete time with infinite horizon, a homogeneous consumption good, and complete markets.

The Model

Arrow-Debreu economy in discrete time with infinite horizon, a homogeneous consumption good, and complete markets.

- $s_t \in \{1, 2, \ldots, S\}$ is the state realized at time t, $\sigma = (s_1, s_2, \ldots, s_t, \ldots)$ is a path, and $\sigma_t = (s_1, s_2, \ldots, s_t)$ is a partial history until time t;

The Model

Arrow-Debreu economy in discrete time with infinite horizon, a homogeneous consumption good, and complete markets.

- $s_t \in \{1, 2, \ldots, S\}$ is the state realized at time t, $\sigma = (s_1, s_2, \ldots, s_t, \ldots)$ is a path, and $\sigma_t = (s_1, s_2, \ldots, s_t)$ is a partial history until time t;
- the economy is populated by N agents with subjective beliefs $p_i(\sigma_t)$, receiving a stream of non-zero and uniformly bounded endowments $(e_i(\sigma_t))_{t=0}^{\infty}$;

The Model

Arrow-Debreu economy in discrete time with infinite horizon, a homogeneous consumption good, and complete markets.

- $s_t \in \{1, 2, \ldots, S\}$ is the state realized at time t, $\sigma = (s_1, s_2, \ldots, s_t, \ldots)$ is a path, and $\sigma_t = (s_1, s_2, \ldots, s_t)$ is a partial history until time t;
- the economy is populated by N agents with subjective beliefs $p_i(\sigma_t)$, receiving a stream of non-zero and uniformly bounded endowments $(e_i(\sigma_t))_{t=0}^{\infty}$;
- each agent maximizes its geometrically discounted expected utility of consumption $c_i(\sigma_t)$.

Problem of the agent, equilibrium, asymptotic outcomes

Given $q(\sigma_t)$, each agent $i = 1, \ldots, N$ solves

$$\max_{\{c_i(\sigma_t),\, \forall t, \sigma\}} \mathrm{E}_{p_i}\left[\sum_{t=0}^{\infty} \beta_i^t u_i(c_i(\sigma_t))\right] \text{ s.t. } \sum_{t=0}^{\infty} \sum_{\sigma_t \in \Sigma_t} q(\sigma_t)\left(e_i(\sigma_t) - c_i(\sigma_t)\right) \geq 0,$$

with $\beta_i \in (0, 1)$ and u^i a continuously differentiable, increasing, strictly concave, and satisfies the Inada condition at zero. At equilibrium,

$$\sum_{i=1}^{N} c_i(\sigma_t) = \sum_{i=1}^{N} e_i(\sigma_t) = e(\sigma_t), \forall \sigma_t.$$

Considering a true probability p and assuming that any p_i is absolutely continuous w. r. t. p, the equilibrium exists.

Problem of the agent, equilibrium, asymptotic outcomes

Given $q(\sigma_t)$, each agent $i = 1, \dots, N$ solves

$$\max_{\{c_i(\sigma_t),\, \forall t,\sigma\}} \mathrm{E}_{p_i}\left[\sum_{t=0}^{\infty} \beta_i^t u_i(c_i(\sigma_t))\right] \text{ s.t. } \sum_{t=0}^{\infty} \sum_{\sigma_t \in \Sigma_t} q(\sigma_t)\,(e_i(\sigma_t) - c_i(\sigma_t)) \geq 0,$$

with $\beta_i \in (0,1)$ and u^i a continuously differentiable, increasing, strictly concave, and satisfies the Inada condition at zero. At equilibrium,

$$\sum_{i=1}^{N} c_i(\sigma_t) = \sum_{i=1}^{N} e_i(\sigma_t) = e(\sigma_t), \forall \sigma_t.$$

Considering a true probability p and assuming that any p_i is absolutely continuous w. r. t. p, the equilibrium exists.

Definition

An agent i *vanishes* if, p-almost surely, $\lim_{t\to\infty} c_i(\sigma_t) = 0$. It *survives* if it does not vanish.

Connecting survival with beliefs

Consider the relative entropy of conditional probabilities and its partial average:

$$D_{p|p_i}(\sigma_t) = \sum_{s=1}^{S} p(s \mid \sigma_t) \log \frac{p(s \mid \sigma_t)}{p_i(s \mid \sigma_t)} \text{ and } \overline{D}_{p|p_i}(\sigma_t) = \frac{1}{t+1} \sum_{\tau=0}^{t} D_{p|p_i}(\sigma_\tau) \, .$$

Connecting survival with beliefs

Consider the relative entropy of conditional probabilities and its partial average:

$$D_{p|p_i}(\sigma_t) = \sum_{s=1}^{S} p(s \mid \sigma_t) \log \frac{p(s \mid \sigma_t)}{p_i(s \mid \sigma_t)} \text{ and } \overline{D}_{p|p_i}(\sigma_t) = \frac{1}{t+1} \sum_{\tau=0}^{t} D_{p|p_i}(\sigma_\tau) \,.$$

Proposition

Given two agents i and j, assume that $\exists L > 0$ such that, p-almost surely, $\| \log p(\cdot \mid \sigma_t)/p_h(\cdot \mid \sigma_t)\|_\infty < L$, $h = i, j$. Then, $\forall \alpha < 1/2$, p-almost surely, for large t,

$$\frac{1}{t} \log \frac{u_i'(c_i(\sigma_t))}{u_j'(c_j(\sigma_t))} = \left(\log \beta_j - \overline{D}_{p|p_j}(\sigma_{t-1})\right) - \left(\log \beta_i - \overline{D}_{p|p_i}(\sigma_{t-1})\right) + o\left(t^{-\alpha}\right) .$$

Moreover if, p-almost surely,

$$\log \beta_j - \log \beta_i + \liminf_{t \to \infty} \left(\overline{D}_{p|p_i}(\sigma_t) - \overline{D}_{p|p_j}(\sigma_t)\right) > 0,$$

then agent i vanishes.

Belief formation

Consider K i.i.d. measures with conditional probabilities $\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_K$ such that $\boldsymbol{\pi}_k = (\pi_k(1), \pi_k(2), \ldots, \pi(S)) \in \Delta_+^{S-1}$; $\exists \epsilon, d\pi > 0$ such that $\pi_k(s) > \epsilon$ and $\|\boldsymbol{\pi}_k - \boldsymbol{\pi}_h\| > d\pi, \forall s, k, h$.

Assumption

The individual conditional probabilities of the agents belong to the convex hull H_K generated by the conditional probabilities of the K models, $\forall \sigma_t$

$$(p_i(1 \mid \sigma_t), \ldots, p_i(S \mid \sigma_t)) \in H_K = \left\{ \sum_{k=1}^{K} \eta_k \boldsymbol{\pi}_k \mid \sum_{k=1}^{K} \eta_k = 1, \eta_k \geq 0 \right\} \subseteq \Delta_+^{S-1}.$$

Moreover, $\exists L > 0$ such that, $\forall k$ and $\forall \sigma_t$, $\| \log p(\cdot \mid \sigma_t)/\pi_k(\cdot) \|_\infty < L$.

Belief formation

Let $w_{i,k}(\sigma_t)$ be the weight agent i attaches to model k after having observed the partial history σ_t. Then, $\forall s$,

$$p_i(s|\sigma_t) = \sum_{k=1}^{K} w_{i,k}(\sigma_t)\,\pi_k(s), \text{ with } w_{i,k}(\sigma_t) \geq 0, \forall k, \text{ and } \sum_{k=1}^{K} w_{i,k}(\sigma_t) = 1.$$

Learning processes differ on how they compute the weights.

Learning processes I

- **Bayesian and under-reaction learning** (Epstein et al., 2010; Massari, 2020):

$$w_{i,k}(\sigma_t) = \lambda_i \, w_{i,k}(\sigma_{t-1}) + (1 - \lambda_i) \frac{\pi_k(s_t) \, w_{i,k}(\sigma_{t-1})}{p_i(s_t|\sigma_{t-1})} \quad \forall k, t, \sigma \, ,$$

with $\lambda_i \in [0, 1)$. Setting $\lambda_i = 0$, Bayesian learning is recovered;

Learning processes II

- **Moving average learning**: the agent takes a reference learning process p^* and applies a moving average of width M_i to the sequence of probabilistic predictions generated by it,

$$p_i(s \mid \sigma_t) = \begin{cases} p^*(s \mid \sigma_t) & \text{if } t < M_i - 1\,, \\ M_i^{-1} \sum_{m=1}^{M_i} p^*(s \mid \sigma_{t-m+1}) & \text{if } t \geq M_i - 1\,. \end{cases}$$

In terms of weights: $w_{i,k}(\sigma_t) = M_i^{-1} \sum_{m=1}^{M_i} w_k^*(\sigma_{t-m+1})$ if $t \geq M_i - 1$.

Learning processes III

- **Limited memory Bayesian learning**:

$$w_{i,k}(\sigma_t) = \frac{\pi_k(s_t) w_{i,k}(\sigma_0)}{\sum_{k'=1}^{K} \pi_{k'}(s_t) w_{i,k'}(\sigma_0)},$$

the agent deliberately forgets observations in the past (extreme short memory case);

General Results

Proposition

Define $\pi^*(\sigma_t) = \max_{k \in \{1,\dots,K\}} \{\pi_k(\sigma_t)\}$. *For any Bayesian agent i and* $\forall \alpha < 1/2$, *p-almost surely, for large t,*

$$\overline{D}_{p|p_i}(\sigma_{t-1}) - \overline{D}_{p|\pi^*(\sigma_t)}(\sigma_{t-1}) = o\left(t^{-\alpha}\right).$$

General Results

Proposition

Define $\pi^*(\sigma_t) = \max_{k\in\{1,\ldots,K\}}\{\pi_k(\sigma_t)\}$*. For any Bayesian agent i and* $\forall\alpha < 1/2$*, p-almost surely, for large t,*

$$\overline{D}_{p|p_i}(\sigma_{t-1}) - \overline{D}_{p|\pi^*(\sigma_t)}(\sigma_{t-1}) = o\left(t^{-\alpha}\right).$$

Proposition

For any underreacting agent i and $\forall\alpha < 1/2$*, it is p-almost surely, for large t,*

$$\overline{D}_{p|p_i}(\sigma_{t-1}) - \overline{D}_{p|\pi_k}(\sigma_{t-1}) \leq o\left(t^{-\alpha}\right), \forall k \in \{1, 2, \ldots, K\}.$$

Parametric misspecification

Assumption

The true measure p is an i.i.d. process whose conditional distribution is described by the vector $\boldsymbol{\pi} = (\pi(1), \pi(2), \ldots, \pi(S)) \in \Delta_+^{S-1}$*, such that* $p(s_t \mid \sigma_{t-1}) = \pi(s_t)$ *and,* $\forall k \in \{1, 2, \ldots, K\}$, $\|\boldsymbol{\pi}_k - \boldsymbol{\pi}\| > 0$.

Parametric misspecification

Assumption

The true measure p is an i.i.d. process whose conditional distribution is described by the vector $\boldsymbol{\pi} = (\pi(1), \pi(2), \dots, \pi(S)) \in \Delta_{+}^{S-1}$*, such that* $p(s_t \mid \sigma_{t-1}) = \pi(s_t)$ *and,* $\forall k \in \{1, 2, \dots, K\}$*,* $\|\boldsymbol{\pi}_k - \boldsymbol{\pi}\| > 0$*.*

Under this Assumption, p-almost surely

$$\lim_{t \to \infty} \overline{D}_{p|\pi_k}(\sigma_t) = D_{\pi|\pi_k} = \sum_{s=1}^{S} \pi(s) \log \frac{\pi(s)}{\pi_k(s)} > 0.$$

Results

Proposition

If $\boldsymbol{\pi} \in H_k$, then for any underreacting agent i and $\forall \alpha < 1/2$, it is p-almost surely, for large t,

$$\overline{D}_{p|p_i}(\sigma_t) \leq \frac{1-\lambda_i}{2(\lambda_i+\epsilon)^2} + \frac{o\left(t^{-\alpha}\right)}{1-\lambda_i}.$$

Results

Proposition

If $\pi \in H_k$, then for any underreacting agent i and $\forall \alpha < 1/2$, it is p-almost surely, for large t,

$$\overline{D}_{p|p_i}(\sigma_t) \leq \frac{1-\lambda_i}{2(\lambda_i+\epsilon)^2} + \frac{o\left(t^{-\alpha}\right)}{1-\lambda_i}.$$

Proposition

For any agent i that uses the limited memory Bayesian learning process, it is $\overline{D}_{p|p_i}(\sigma) = \lim_{t\to\infty} \overline{D}_{p|p_i}(\sigma_t) = \sum_{s=1}^{S} \pi(s) D_{p|p_i}(s)$ p-almost surely.

Results

Proposition

Given a moving average learning process p_i *with the reference learning process* p^**, consider* $\sigma^2(s,\sigma_t) = \sum_{m=0}^{M-1} \left(p^*(s \mid \sigma_{t-m}) - p_i(s,\sigma_t)\right)^2 / M$*. Then,*

$$\frac{\sigma^2(\sigma_t)}{2(1-\epsilon)} \leq \frac{1}{M}\sum_{m=0}^{M-1} D_{p|p^*}(\sigma_{t-m}) - D_{p|p_i}(\sigma_t) \leq \frac{\sigma^2(\sigma_t)}{2\epsilon}.$$

If $\overline{D}_{p|p^*}(\sigma) = \lim_{t\to\infty} \overline{D}_{p|p^*}(\sigma_t)$ *exists, then* $\limsup_{t\to\infty} \overline{D}_{p|p_i}(\sigma_t) \leq \overline{D}_{p|p^*}(\sigma)$*, with strict inequality if* $\exists \varepsilon > 0$ *such that* $\sigma^2(\sigma_t) > \varepsilon$*.*

Example

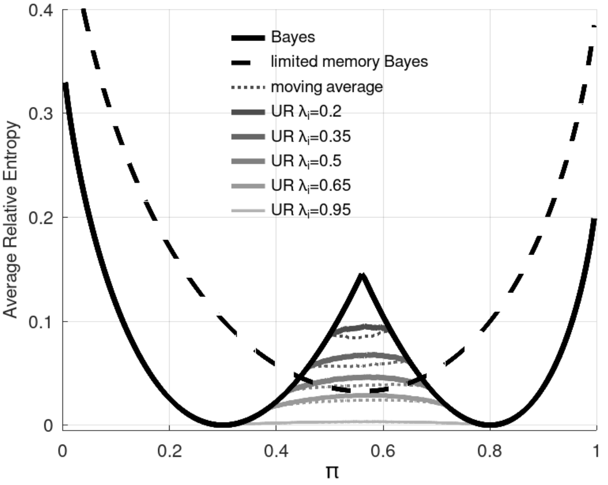

Figure: $S = K = 2$, $\pi_1 = 0.3$, $\pi_2 = 0.8$, $M_i = 10$, π is the true probability of state 1.

Structural misspecification

Assumption

The true measure p follows a discrete-time Markov chain with transition matrix P: $p(s_{t+1}|\sigma_t) = P_{s_t,s_{t+1}}\ \forall t, \sigma$ *and* $p(s|\sigma_0) = p_{s,0}$ *with* $p_{s,0} > 0\ \forall s \in \{1, 2, \ldots, S\}$*. For any* $(s, s') \in \{1, 2, \ldots, S\} \times \{1, 2, \ldots, S\}$*,* $P_{s,s'} > 0$*.*

Structural misspecification

Assumption

The true measure p follows a discrete-time Markov chain with transition matrix P: $p(s_{t+1}|\sigma_t) = P_{s_t,s_{t+1}}\ \forall t,\sigma$ and $p(s|\sigma_0) = p_{s,0}$ with $p_{s,0} > 0\ \forall s \in \{1,2,\dots,S\}$. For any $(s,s') \in \{1,2,\dots,S\} \times \{1,2,\dots,S\}$, $P_{s,s'} > 0$.

Proposition

For any $k = 1,\dots,K$, p-almost surely,

$$\lim_{t\to\infty} \overline{D}_{p|\pi_k}(\sigma_t) = \overline{D}_{p|\pi_k}(\sigma) = \overline{D}_{\pi|\pi_k}(\sigma) + \overline{D}_{p|\pi}(\sigma),$$

where

$$\overline{D}_{\pi|\pi_k}(\sigma) = \sum_{s=1}^{S} \pi(s) \log \frac{\pi(s)}{\pi_k(s)} \text{ and } \overline{D}_{p|\pi}(\sigma) = \sum_{s'=1}^{S} \pi(s') \sum_{s=1}^{S} P_{s',s} \log \frac{P_{s',s}}{\pi(s)}.$$

Example

$K = S = 2$, $\boldsymbol{\pi}_1 = (\pi_1, 1 - \pi_1)$, $\boldsymbol{\pi}_2 = (\pi_2, 1 - \pi_2)$, and states of nature appear according to a Markov chain with transition matrix

$$P = \begin{bmatrix} P_{1,1} & 1 - P_{1,1} \\ P_{2,1} & 1 - P_{2,1} \end{bmatrix} .$$

The invariant distribution reads

$$\boldsymbol{\pi} = \left(\frac{P_{2,1}}{1 - P_{1,1} + P_{2,1}} , \frac{1 - P_{1,1}}{1 - P_{1,1} + P_{2,1}} \right) .$$

Example:

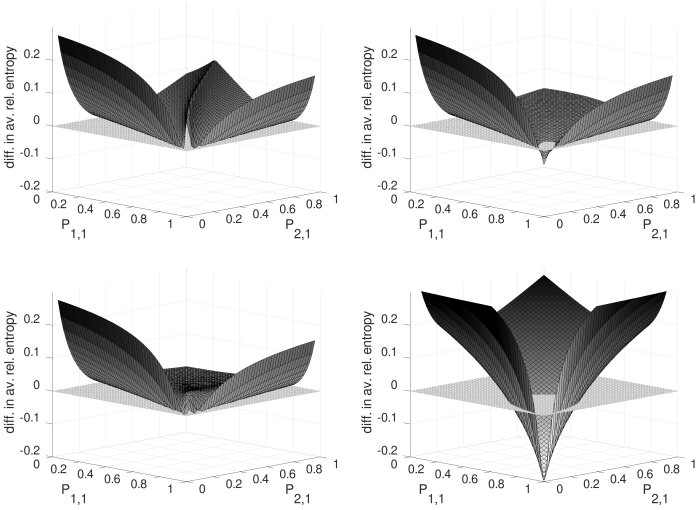

Figure: Diff. in av. rel. entropy btw B, UR $\lambda = 0.65$, MA $M = 20$ on UR, LMB and invariant.

Conclusions

- Ranking learning processes with respect to their survival prospects (and identifying general survival mechanisms) under model misspecification is hard;

Conclusions

- Ranking learning processes with respect to their survival prospects (and identifying general survival mechanisms) under model misspecification is hard;
- Under parametric misspecification, learning processes build upon a smoothing approach have a selection advantage over generic regions of the parameter space;

Conclusions

- Ranking learning processes with respect to their survival prospects (and identifying general survival mechanisms) under model misspecification is hard;
- Under parametric misspecification, learning processes build upon a smoothing approach have a selection advantage over generic regions of the parameter space;
- The advantage partially disappears when structural misspecification is considered: a trade off between approximating the projection of the true model on the space on which the agents learn and adapting to the part of the true model that cannot be represented in that space emerges;

Conclusions

- Ranking learning processes with respect to their survival prospects (and identifying general survival mechanisms) under model misspecification is hard;
- Under parametric misspecification, learning processes build upon a smoothing approach have a selection advantage over generic regions of the parameter space;
- The advantage partially disappears when structural misspecification is considered: a trade off between approximating the projection of the true model on the space on which the agents learn and adapting to the part of the true model that cannot be represented in that space emerges;
- The **ecology of traders** and **type of model misspecification** matter for understanding selection outcomes and, thus, long-run asset valuation.

References I

Antico, A., G. Bottazzi, and D. Giachini (2023). On the evolutionary stability of the sentiment investor. In D. Bourghelle, P. Grandin, F. Jawadi, and P. Rozin (Eds.), *Behavioral Finance and Asset Prices: The Influence of Investor's Emotions*, pp. 155–173. Cham: Springer International Publishing.

Box, G. E. (1976). Science and statistics. *Journal of the American Statistical Association 71*(356), 791–799.

Cerreia-Vioglio, S., L. P. Hansen, F. Maccheroni, and M. Marinacci (2020). Making decisions under model misspecification. *arXiv preprint arXiv:2008.01071*.

Cox, D. R. (1995). Comment on "model uncertainty, data mining and statistical inference". *Journal of the Royal Statistical Society, Series A 158*, 455–456.

Epstein, L. G., J. Noor, and A. Sandroni (2010). Non-bayesian learning. *The BE Journal of Theoretical Economics 10*(1), 1–20.

Fudenberg, D., G. Romanyuk, and P. Strack (2017). Active learning with a misspecified prior. *Theoretical Economics 12*(3), 1155–1189.

References II

Hansen, L. P. (2014). Nobel lecture: Uncertainty outside and inside economic models. *Journal of Political Economy 122*(5), 945–987.

Hansen, L. P. and T. J. Sargent (2022). Structured ambiguity and model misspecification. *Journal of Economic Theory 199*, 105165.

Marinacci, M. and F. Massari (2019). Learning from ambiguous and misspecified models. *Journal of Mathematical Economics 84*, 144–149.

Massari, F. (2020). Under-reaction: Irrational behavior or robust response to model misspecification? *Available at SSRN 3636136*.

Massari, F. (2021). Price probabilities: A class of bayesian and non-bayesian prediction rules. *Economic Theory 72*, 133–166.