Work after 9 Nov 2019

Form the FEs correctly:

$$f_t^m = \hat{E}_t(\pi_{t+1} \mid \mathcal{I}_t^m) = \hat{E}_t(\pi_{t+1} \mid s_t, \bar{\pi}_{t-1})$$

$$f_t^e = \hat{E}_t(\pi_{t+1} \mid \mathcal{I}_t^e) = \hat{E}_t(\pi_{t+1} \mid s_t, \bar{\pi}_t)$$

$$\left.\begin{aligned} FE_t^m &= \pi_{t+1} - f_t^m \\ FE_t^e &= \pi_{t+1} - f_t^e \end{aligned}\right\}$$

So in either case I just need to subtract F_t from $ysim_{t+1}$.

and note that these FE_t are realized at $t+1$.

The problem is that the FEs I construct this way aren't equal to the ones I get from the sim_learn.m code. This is puzzling b/c in principle they come from the same simulated π – same fcst of π.

The problem is that the FE coming out of sim_learn

1) always changes, despite IRF-ing & averaging
2) There are diffs btwn $FE^{shocked}$ & $FE^{unshocked}$ even before I impose the shock δ (!)

FEs are solved. 14 Nov 2019

I think that the criss-crossing of fcsts is well-understood: when cgain, you update your fcst too much and so your FE switches sign and volá.
At a certain point, your FE is small enough so that no overupdating of expectations happens any more.
→ my bet is you can kill this overupdating / crisscrossing w/ a sufficiently low gain
yes, w/ $\bar{g} = 0.1$ (instead of 0.145) you already have dgain & cgain similar at $t=5$
w/ $\bar{g} = 0.0145$ they're identical at $t=25$ too

- But you always get some overshooting, whether it's in the 2nd period (cgain) or later on (dgain)
- Moreover, it's puzzling that $i \uparrow$ as $\pi < 0$ in 2nd period

One way to get perfectly normal, RE-like responses is to set $\alpha = 1$ b/c then $f_a \approx f_b$. But even $\alpha = 0.99$ gets a quite sig diff btwn f_a & f_b & overshooting too!

$$fa = \frac{a}{1-\alpha\beta} + b\left(I_{nx} - \alpha\beta hx\right)^{-1} s$$

What is $\frac{1}{1-\alpha\beta} = 50.2513$ and $\frac{1}{1-\beta} = 100$

for $\alpha = \beta = 0.99$?

$$\left(I_4 - \alpha\beta hx\right)^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2.4275 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0.0049 & 2.5069 & 1.4618 & 1 \end{bmatrix}$$

and $$\left(I_4 - \beta hx\right)^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2.4631 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0.0050 & 2.3643 & 1.4776 & 1 \end{bmatrix}$$

ha! The diff in fa & fb is most pronounced in the part that comes from the intercept!

I think this would change a bit as a moves away from 1 (PTO)

$\alpha = 0.5$, $\beta = 0.99$

$$\frac{1}{1-\alpha\beta} = 1.9802 \quad \text{and} \quad \frac{1}{1-\beta} = 100$$

$$(I_4 - \alpha\beta h_x)^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1.4225 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0.0025 & 0.6827 & 0.7388 & 1 \end{bmatrix}$$

$$\text{and } (I_4 - \beta h_x)^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2.4631 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0.0050 & 2.3643 & 1.9776 & 1 \end{bmatrix}$$

Yes, now the part relating the slope is diff, but the part relating the intercept is even more diff!

$\Rightarrow$ <u>the lower α (the higher κ, the less price stickiness)</u>
<u>the more f_0 loads on the intercept both in absolute terms</u>
<u>(f_0 reacts more) and relative terms (vs. the slope)</u>

$\Rightarrow$ this may be driving (some of) the overshooting b/c for the std param value of $\alpha = 0.5$, f_b is almost 50 times more driven by the intercept than the slope + shocks. $\Rightarrow$ so overreaction in updating the intercept drives f_b, which is what drives x_t up for again.

Let's interpret

$$\alpha\beta \cdot hx \qquad \text{vs} \qquad \beta \cdot hx$$

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.297 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0.0025 & 0.48 & 0.7388 & 0 \end{bmatrix} \quad \begin{array}{c} r^n \\ \bar{i} \\ u \\ i_{t-1} \end{array} \quad \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.594 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0.005 & 0.96 & 1.4776 & 0 \end{bmatrix}$$

2 differences

1) effect of shocks on i_t
2) effect of $\bar{i}$ on s

$\left.\begin{array}{l} \text{1)} \\ \text{2)} \end{array}\right\}$ $\alpha < 1$ mutes these somewhat

$\rightarrow$ you're just discounting shocks in the future more!

Also, when $\alpha = 1$, then $f_a(1)$ doesn't matter for (x, π) " $E\pi$

The puzzling $i\uparrow$ at $t=2$

$A_a(3,1) = 0.5928 \rightarrow i\uparrow$ if $f_a(1)$ " $E\pi$ $\uparrow$

$A_b(3,1) = -0.0978 \rightarrow i\downarrow$ if $f_b(1)\uparrow$

$\Rightarrow$ so when f_b moves a lot more than f_a, (which is in general not true for avg dgain), then $i\uparrow$ even when $E[\pi]$ is $\downarrow$ (!)

But why?

if $\psi_x\uparrow$ (now it's 0) then $A_a(3,1)\downarrow$ and

$A_b(3,1)\downarrow$ too!

But $A_a(3,1)$ never <0, not even for $\psi_x = 5$.

when $\psi_x = 0$, it's b/c $\pi\downarrow$ when $f_b\uparrow$

$\rightarrow$ it seems like $i\uparrow$ in $t=2$ b/c π is $\uparrow$ from $t=1$ to $t=2$

:S

$$Aa(3,1) = \psi_\pi Aa(1,1) + \underbrace{\psi_x Aa(2,1)}_{=0}$$

$$\underbrace{Ab(3,1) = \psi_\pi Ab(1,1)}_{\text{these are true!}}$$

$\Rightarrow$ ah I see: $i\uparrow$ at $t=2$ b/c it was going up much more at $t=1$ due to the innovation, but since π fell so much, this depressed i a lot. At $t=2$, since $\pi\uparrow$ (but is still <0), i is depressed below $0.6 \cdot 1^{\text{"}\delta}$ (it's only ≈ 0.1) but it's not depressed as much

<u>Puzzling i-response</u>

$$i_t = \underset{\downarrow}{\pi} + \underset{\uparrow}{\text{innovation } (\delta)}$$

Initially $|\delta| > |\pi|$

At $t=2$ $|\pi|$ shrinks so $i\uparrow$

What remains to be understood is why the overshooting happens regardless, just later:

- maybe whats going on is that $\mathbb{E}[\bar{\pi}]$ are pushing stuff up but $i\uparrow$ is pushing them down, and i recedes faster

Check: if $\bar{i}$-shock is iid, overshooting should happen at $t=2$

$\Rightarrow$ exactly, and it does!

The only thing that isn't a 100% clear is why the $\ominus$ reaction to expectations, when RE doesn't have this?

In RE: $x_t = E_t x_{t+1} - \sigma E_t(i_t - \pi_{t+1})$

$$\pi_t = \kappa x_t + \beta E_t \pi_{t+1}$$

$$i_t = \psi_\pi \pi_t$$

$$x_t = -\sigma i_t + E_t x_{t+1} + \sigma E_t \pi_{t+1}$$

$$\pi_t = \kappa x_t + \beta E_t \pi_{t+1}$$

$$x_t = -\sigma \psi_\pi \pi_t + E_t x_{t+1} + \sigma E_t \pi_{t+1}$$

$$x_t = -\sigma \psi_\pi \left[\kappa x_t + \beta E_t \pi_{t+1}\right] + E_t x_{t+1} + \sigma E_t \pi_{t+1}$$

$$x_t = -\sigma\psi_\pi\left[\kappa x_t + \beta E_t\pi_{t+1}\right] + E_t x_{t+1} + \sigma E_t\pi_{t+1}$$

$$(1+\sigma\psi_\pi\kappa)x_t = -\sigma\psi_\pi\beta E_t\pi_{t+1} + E_t x_{t+1} + \sigma E_t\pi_{t+1}$$

$$x_t = \underbrace{\frac{1}{w}\sigma(1-\psi_\pi\beta)}_{<0\ (!)} E_t\pi_{t+1} + \frac{1}{w}E_t x_{t+1}$$

$\Rightarrow$ RE has it too, only $E[\cdot]$ don't move as much

$$\Rightarrow \pi_t = \frac{\kappa}{w}\sigma(1-\psi_\pi\beta)E_t\pi_{t+1} + \frac{\kappa}{w}E_t x_{t+1} + \beta E_t\pi_{t+1}$$

$$\pi_t = \underbrace{\left[\frac{\kappa}{w}\sigma(1-\beta\psi_\pi)+\beta\right]}_{\text{likely }>0}E_t\pi_{t+1} + \frac{\kappa}{w}E_t x_{t+1}$$

(for current params = 0.9298)

RE:

$$x_t = \ominus E_t\pi_{t+1} + \oplus E_t x_{t+1}$$

$$\pi_t = \oplus E_t\pi_{t+1} + \oplus E_t x_{t+1}$$

Learning

$$x_t = \ominus E_t\pi^{fa}_{t+1} + \ominus E_t\pi^{fb}_{t+1} + \ominus E_t x^{fa}_{t+1} + \oplus E_t x^{fb}_{t+1}$$

$$\pi_t = \oplus E_t\pi^{fa}_{t+1} + \ominus E_t\pi^{fb}_{t+1} + \oplus E_t x^{fa}_{t+1} + \oplus E_t x^{fb}_{t+1}$$

$$x_t = \ominus E\pi + \ominus EX \qquad (1)$$

this is not a mistake

$$\pi_t = \oplus E\pi + \oplus EX \qquad (2)$$

① $\frac{-b\psi_\pi}{w} k\alpha\beta + \frac{1}{w}(1-\beta) = \frac{1-\beta - b\psi_\pi k\alpha\beta}{w} < 0$

② $\left(1 - \frac{kb\psi_\pi}{w}\right)(1-\alpha)\beta + \frac{k}{w} b(1-\beta\psi_\pi) > 0$

⟹ why do we have this diff btwn RE & Learn? 15 Nov 2019

$b\psi_\pi k\alpha\beta \overset{!}{<} 1-\beta$

$b\psi_\pi k\alpha\beta + \beta < 1$

$(b\psi_\pi k\alpha + 1)\beta < 1$ but it's 1.1150.

In the RE world, x depends on E(x) only directly

My conjecture is that $E(\pi)$ in RE will incorporate $E(x)$ in some way. So π must depend stronger on $E(\pi)$ in RE than in learning.

RE: π on $E(\pi)$: $\frac{kb + \beta}{b\psi_\pi} = 0.7722$ under current params

Learn: $\left(1 - \frac{kb\psi_\pi}{w}\right)(1-\alpha)\beta + \frac{k}{w} b(1-\beta\psi_\pi) = 0.33$ —"—

→ So yes, this is true

In fact, you can reason that in RE,

$\pi = \underset{\oplus}{E(\pi)}$ only b/c only via

while in learning

$\pi = \underset{\oplus}{E(\pi)}$ but part of this is $\oplus \rightarrow$ from f_a

$\ominus \rightarrow$ from f_b

<u>Ryan meeting</u> <u>15 Nov 2019</u>

fix point where find the gain that min Var(FE)

data is generated by a gain = 0.145 gain,

given this, let an agent set a best gain

→ it must be lower than 0.145 !

Analog to RE for the gain param.

↳ Pooya Molavi's JMP does this, using a Kullback - Leibler distance.

Stage 0: Establish that again learning causes excess volatility.

0.1: Do learning rule where I don't do RE-part
$PLM = \bar{\pi}_{t-1}$ and that's it.

0.2: Try learning the slope.
$\Rightarrow$ in those contexts, do I continue to get the flipflops?

2. Didn't quite get to the bottom of RE vs. learning loading on $E(\cdot)$
$\rightarrow$ connect to those equations

3. These features can become worse if
$\uparrow \psi_\pi$.

Do it in a week. Schedule to talk to Basu after.
Tell him: in learning models, there's this endemic instability. This can become worse if $\psi_\pi \uparrow$.
Here's how it works.

Work after

15 Nov: did "only-mean" PLM and "slope & constant"

✓ • check that the latter is correct

✓ • print figs w/o cutting them off

• polish explanation of $E(\cdot) \to z$

• do the fixed point thing, use Molavi.

Reading Molavi JMP 16 Nov 2019

I'm not superimpressed b/c the "constrained RE eqb" (CREE) is really just saying that give agents a set of models Θ and let them choose (their expectation formation) the subset Θ^* that min $H(\cdot)$ where $H(\theta, T)$ is the Kullback-Leibler distance between model θ and the ALM T. (Sometimes Θ^* is a singleton, Molavi calls this a pure CREE.)

He shows (thm 2 & 3) that Bayesian & adaptive learning coincide w/ a CREE in the LR

→ of course b/c the CREE must be = REE if the REE $\in \Theta$! This is why Molavi says that the "LR behavior of the econ is independent of ... the learning process" b/c they all converge to REE! (Unless they are not E-stable, which is the analog of Molavi's concept of REE $\notin \Theta$, i.e. when agents don't include the REE in the set of models they consider)

→ ok so trying to solve for $\bar{g}^*$

$\bar{g}^* = \arg\min FEV$

I have 2 ways to construct FEV.

1) analytically (don't know if possible)

2) numerically in Matlab

↳ here I'm confused whether the FEV is across time or cross-section ← I suppose that.

$Var(X) = E(X^2) - E(X)^2$

If $X = FE$, then $E(X) = 0$, so $FEV = E(FE^2)$

$$\text{var}(X) = E[\underbrace{(X - E[X])}_{FE}^2]$$

this is why

$FEV(x) = \text{var}(X)$ when you initialize!

Otherwise $FEV = E\left[(X_{t+k} - X_{t+k,t})(X_{t+k} - X_{t+k,t})'\right]$

fcst

Ok let's clarify one thing:

$$\underbrace{\text{Var}(X) = E[(X - E[X])^2]}_{} = E(X^2) - E(X)^2$$

it seems right now that $\text{var}(X) = FEV(X)$ are the same thing?!

Leaving that aside for a moment

- my fcsts are given by the $PLM(\bar{g})$
- $FE_{t-1} = \pi_t - PLM^e_{t-1}(\bar{g})$
- $FEV_{t-1} = E\left[(\pi_t - PLM^e_{t-1}(\bar{g}))^2\right]$

Let's take a general case: $PLM^e_{t-1}(\bar{g}) = \phi_{t-1}\begin{bmatrix} 1 \\ s_{t-1} \end{bmatrix}$

$$\Rightarrow FEV_{t-1} = E\left[\left(\pi_t - \phi_{t-1}(\bar{g})\begin{bmatrix} 1 \\ s_{t-1} \end{bmatrix}\right)^2\right]$$

$$\phi_{t-1}(\bar{g}) = \left(\phi'_{t-2} + \bar{g}\left(\pi_t - \phi_{t-2}\begin{bmatrix} 1 \\ s_{t-2} \end{bmatrix}\right)\right)'$$

I don't think I can solve the problem analytically b/c it's so recursive: to find $\bar{g}_t^*$, I need to have $\bar{g}_{t-1}^*$ etc.

Also I'm not sure if I should

- restrict agents to use the same $\bar{g}$ in every period
- make them optimize over $\bar{g}$ in every period.

ok - what I have now is FEV across time → I make 'em min FEV for each history n

→ this gives me N $\bar{g}^*$'s, which I then average.

So far I got 0.00021076 ($2.1076 \cdot 10^{-4}$) ≈ 0.0002

⇒ think more on this tomorrow!

RE vs learning: responses to $E(\cdot)$ 17 Nov 2019

RE:

$$x_t = -\sigma\psi_\pi\left(\kappa x_t + \beta E_t \pi_{t+1}\right) + E_t x_{t+1} + \sigma E_t \pi_{t+1}$$

$$= -\sigma\psi_\pi \kappa x_t - \sigma\beta\psi_\pi E_t \pi_{t+1} + E_t x_{t+1} + \sigma E_t \pi_{t+1}$$

$$(1+\sigma\psi_\pi\kappa) x_t = \sigma(1-\beta\psi_\pi) E_t \pi_{t+1} + E_t x_{t+1}$$

$$x_t = \frac{\sigma(1-\beta\psi_\pi)}{1+\sigma\psi_\pi\kappa} E_t \pi_{t+1} + \frac{1}{1+\sigma\psi_\pi\kappa} E_t x_{t+1}$$

$$\pi_t = \kappa \frac{\sigma(1-\beta\psi_\pi)}{1+\sigma\psi_\pi\kappa} E_t \pi_{t+1} \quad \frac{\kappa}{1+\sigma\psi_\pi\kappa} E_t x_{t+1} + \beta E_t \pi_{t+1}$$

$$\pi_t = \left(\frac{\kappa\sigma(1-\beta\psi_\pi)}{1+\sigma\psi_\pi\kappa} + \beta\right) E_t \pi_{t+1} + \frac{\kappa}{1+\sigma\psi_\pi\kappa} E_t x_{t+1}$$

Learning

$$x_t = \left(-\frac{\sigma\psi_\pi}{w}(1-\alpha)\beta + \frac{\sigma(1-\beta\psi_\pi)}{w}\right) E_t^{\alpha,\beta} \pi_\infty$$

$$\left(-\frac{\sigma\psi_\pi}{w}(\kappa\alpha\beta) + \frac{1-\beta}{w}\right) E_t^{\alpha,\beta} x_\infty$$

$$\pi_t = \left(\left(1 - \frac{\kappa\sigma\psi_\pi}{w}\right)(1-\alpha)\beta + \frac{\kappa\sigma(1-\beta\psi_\pi)}{w}\right) E_t^{\alpha,\beta} \pi_\infty$$

$$+ \left(\left(1 - \frac{\kappa\sigma\psi_\pi}{w}\right)(\kappa\alpha\beta) + \frac{\kappa(1-\beta)}{w}\right) E_t^{\alpha,\beta} x_\infty$$

materials10 → parameter values:

- $\frac{k\tau(1-\beta\psi_\pi)}{1+\tau\psi_\pi k} + \beta = \frac{k\tau \cancel{-k\tau\beta\psi} + \beta \cancel{+\beta\tau\psi_\pi k}}{1+\tau\psi_\pi k} > 0$

- $-\frac{\tau\psi_\pi k\alpha\beta}{1+\tau\psi_\pi k} + \frac{1-\beta}{1+\tau\psi_\pi k} \propto 1-\beta-\tau\psi_\pi k\alpha\beta$

  For this to be positive, we need $\beta + \tau\psi_\pi k\alpha\beta < 1$

  For current params, this is $1.155 > 1$.

- $\left(1-\frac{k\tau\psi_\pi}{1+k\tau\psi_\pi}\right)(1-\alpha)\beta + \frac{k\tau(1-\beta\psi_\pi)}{1+\tau\psi_\pi k}$

$\propto (1+\cancel{k\tau\psi_\pi}-\cancel{k\tau\psi_\pi})(1-\alpha)\beta + k\tau - k\tau\beta\psi_\pi$

$= (1+\alpha)\beta + k\tau(1-\beta\psi_\pi) = 1.4034 > 0$

$\underbrace{\beta+\alpha\beta}_{\approx 1.5} + \underbrace{k\tau}_{\approx 0} - k\tau\underbrace{\beta\psi_\pi}_{\approx\psi_\pi} > 0$ even if $\psi_\pi = 5$

Ok, so now explain why, if I recursively substi into the RE system, why do I not get the learning system? Even though you can pull out the next term from the learning system to reduce to RE.

→ it seems that LIE holds for the idiosyncratic expectation $\hat{E}^i_t \hat{E}^i_{t+1} = \hat{E}^i_t$ (in fact, this IS anticipated utility!) but not for the average expectation: $\hat{E}_t \hat{E}_{t+1} \neq \hat{E}_t$ ⇒ it's a little bit like the distinction btw PLM & ALM b/c indiv act based on $\hat{E}^i_t \hat{E}^i_{t+1} = \hat{E}^i_t$, i.e. thinking that LIE holds, but in the actual law of motion $\hat{E}_t \hat{E}_{t+1}$ turns out not to equal $\hat{E}_t$ since updating happens!

19 Nov 2019

For Susanto, use

- from materials 10: "A more concise rephrasing"
- I think I wanna show IRFs from Dgain & cgain against RE for std params for the 3 shocks (take iid shocks ∀ except monpol.

Ryan meeting

20 Nov 2015

(500 years)

→ do for FEV-min $T = 5 \cdot 400$

→ question: maybe distrib isn't ergodic ⇒ dispersion in $\bar{g}^*$ is too large

Fabio Milani & Preston } have estimated gains

so if I generate a data sequence from RE, and I allow agents to choose gain, optimal is 0.

Ryan conjectures:

"If you do $T = 5 \cdot 400$ → will you squeeze the dispersion and shrink the mean? Yes."

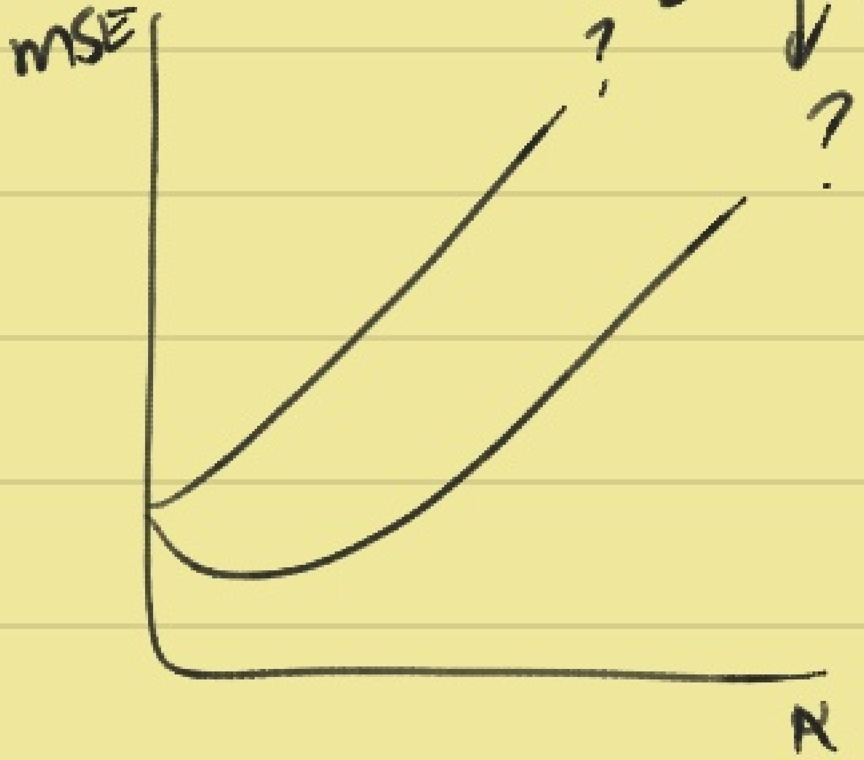
→ But if this exercise gives $\bar{g}^* = 0$ then it's not the right model-based notion of what the $\bar{g}$ should be.

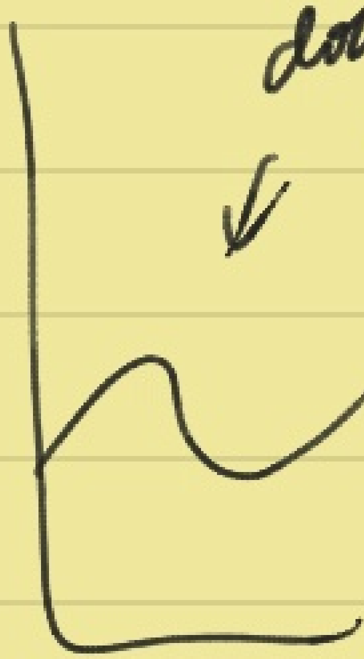
⇒ Ryan says something like: our exercise here is to get a model-based notion of what the gain should be. But we do not yet know what the right notion is.

→ maybe b/c that's saying that the learning model is not optimal. (?)

Plot this

We don't know yet what this fct looks like





does it wobble a lot?

if yes, taking average is necessary but Ryan thinks it won't.

However it could be that this point is a local, but not a global min!

→ allow slightly neg. α to see if it comes back up. That is, if the MSE function looks like this:

If not minimized for reasonable k,

can $\pi - E(\cdot)$ look anchored?

↳ What he means by this is the following:

We have found that the "puzzling IRFs" ("overshooting") is an endemic feature of learning. It decreases when gains are smaller. So we are trying to find a way to pick a reasonable gain k – either from a model & optimality perspective or from the data. And the question is: supp. we have a reasonable k, k^R.

1.) Do we get overshooting for that?

The point is that the IRFs of the econ for k^R are going to be the model's prediction for what unanchored expectations look like.

If overshooting is endemic for k^R, then for anchoring to be a good model, you'd need the overshooting IRFs to fit data.

⇒ So, as Ryan said, this can put me in a dilemma, or, I could call it a crossroads: maybe the anchored E model isn't a good model of $E(\cdot)$? Maybe they have to learn about x and/or i too, or maybe something entirely different...