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 voilà. 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 $fa \approx fb$. But even $\alpha = 0.99$ gets a quite sig diff btwn fa & fb & 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 $\alpha = \beta = 0.99$?

$$\left(I_4 - \alpha\beta hx\right)^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2.4275 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0.0049 & 2.3069 & 1.4678 & 1 \end{bmatrix}$$

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 α moves away from 1 (PTO)

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

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

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

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!

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

⟹ 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 α gain.

Let's interpret

$$
\alpha\beta \cdot h_x \quad \text{vs} \quad \beta \cdot h_x
$$

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

2 differences

1) effect of shocks on i_t $\Big\}$ $\alpha < 1$ mutes these somewhat

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

$\rightarrow$ 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 $f_a(1) \uparrow$ " $E\pi$

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

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

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

⇒ ah I see: i↑ 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 π↑ (but is still <0), i is depressed below $0.6 \cdot 1^{\delta}$ (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↑

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

- maybe whats going on is that $E[\pi]$ 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[·] 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 btw RE & Learn? <u>15 Nov 2019</u>

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

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

$(b\psi_\pi k\alpha + 1)\beta < 1$ 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.
=> 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(·)
→ connect to those equations

3. These features can become worse if I ↑ ψ_{π}.

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 ψ_{π}↑.
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 $\bar{g}^*$

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

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 the.

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

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

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

This is why

$FEV(x) = Var(X)$ when you initialize!

Otherwise $FEV = E[(X_{t+k} - \underset{\uparrow \atop \text{fcst}}{X_{t+k,t}})(X_{t+k} - X_{t+k,t})']$

Ok let's clarify one thing:

$$\underbrace{Var(X) = E[(X - E[X])^2]}_{\text{it seems right now that } Var(X) = FEV(X)} = E(X^2) - E(X)^2$$

are the same thing?!

Leaving that aside for a moment

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

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

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

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

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 $(2.1076 \cdot 10^{-4}) \approx 0.0002$

⇒ 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:

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

- $-\frac{\sigma\psi_\pi\kappa\alpha\beta}{1+\sigma\psi_\pi\kappa} + \frac{1-\beta}{1+\sigma\psi_\pi\kappa} \propto 1-\beta-\sigma\psi_\pi\kappa\alpha\beta$

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

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

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

$\propto (1+\cancel{\kappa\sigma\psi_\pi} - \cancel{\kappa\sigma\psi_\pi})(1-\alpha)\beta + \kappa\sigma - \kappa\sigma\beta\psi_\pi$

$= (1+\alpha)\beta + \kappa\sigma(1-\beta\psi_\pi) = 1.4034 > 0$

$\underbrace{\beta+\alpha\beta}_{\approx 1.5} + \underbrace{\kappa\sigma}_{\approx 0} - \kappa\sigma\underbrace{\beta\psi_\pi}_{\approx\psi_\pi} > 0$ 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 btwn PLM & ALM b/c indis 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

(500 years) 20 Nov 2015

→ 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.

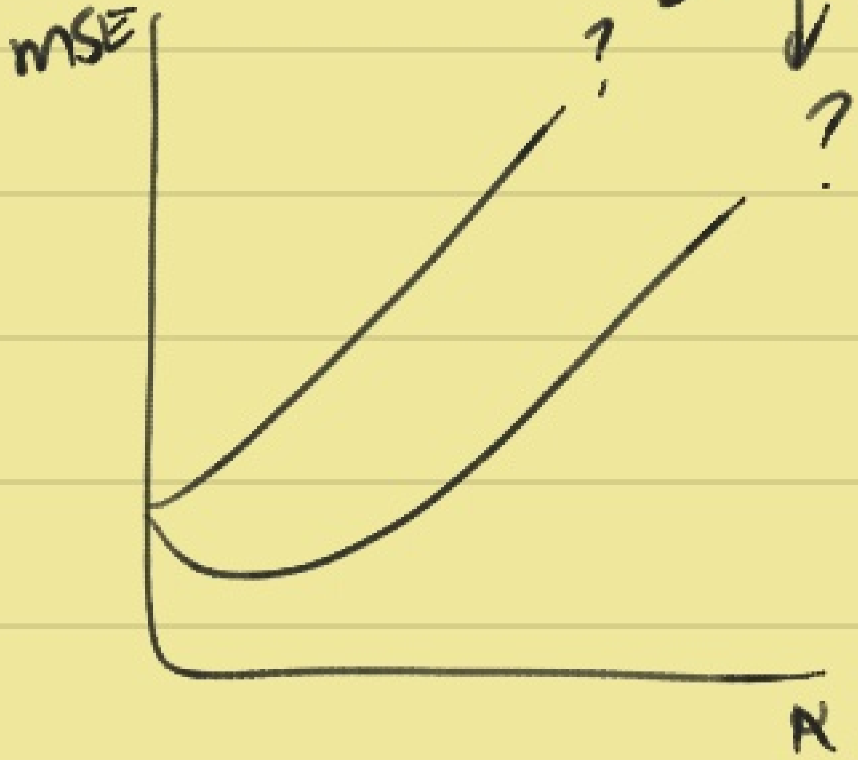
⇒ Ryan says something like: our exercise here is to get a model-based notion of what the gain should be. But we do not yet know what the right notion is.

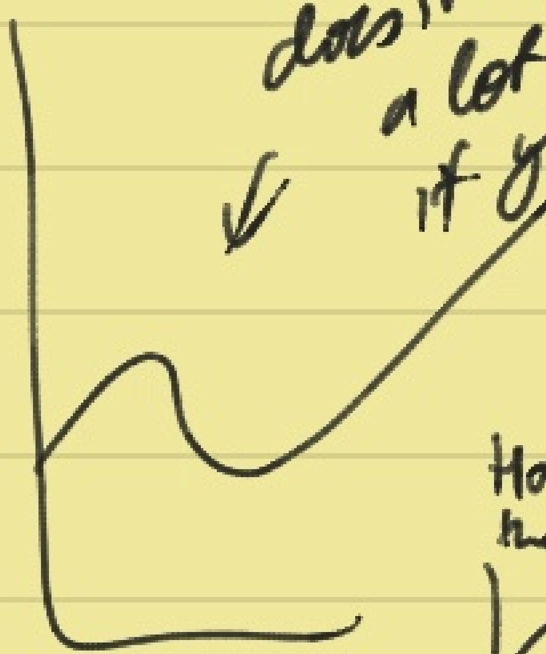
→ maybe b/c that's saying that the learning model is not optimal.

(?)

Plot this

We don't know yet what this fct looks like



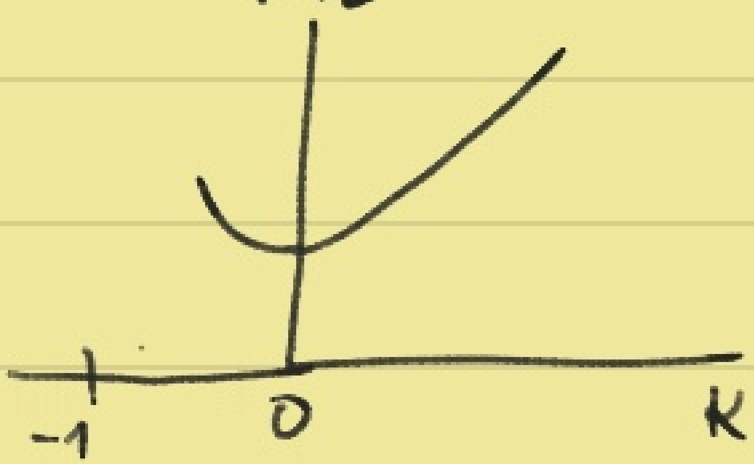


does it wobble a lot?

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. κ 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...

Work after 20 Nov 2019

Let's gather some reasonable numbers for the gain.

Estimates	Calibration	
	0.002	Eusepi & Preston (2011)
$g < 2(1-\beta)$ for beliefs to be stable. For $\beta = 0.99$, $g < 0.02$. $\hat{g} = 0.05$		Eusepi, Giannoni, Preston (2019) Limits
0.145		CEMP
0.0183		Milani (2007)
	0.1, 0.05, 0.03	Williams (2003)
0.062		Branch & Evans (2006)
	0.075, 0.05, 0.025	Orphanides & Williams (2004)
	0.02	Orphanides & Williams (2005)

Avg = 0.05

Avg w/o CEMP & avg estimate w/o CEMP = 0.04

Let's interpret the gain number:

- Eusepi & Preston (2011)

  Data T quarters old receives the weight

  $(1-k)^T$

- Milani (2007)

  $\frac{1}{\text{gain}} \approx$ "how many past observations agents use to form expectations"

Eusepi & Preston (2011) → 500 quarters (⇒ 125 years!)
avg w/o CEMP → 24 quarters (⇒ 6 years)
CEMP → 7 quarters (< 2 years!)

I think it's reasonable that humans shouldn't use more than approx 50 years of data on avg (200 quarters)

↳ $\min(\bar{g}) = 0.005$

They also shouldn't use less than 5 years (20 quarters)

↳ $\max(\bar{g}) = 0.05$

$\bar{g} \in [0.005, 0.05]$, and in particular $0.02 \to 12$ years seems reasonable.

When burnin = 0, T = 400, N = 100 (48 sec)

$\bar{g}^* = 0.0005$ $Var(\bar{g}^*_n) = 1.2602 \cdot 10^{-6}$

→ but it varies per run!

burnin = 50, T = 400, N = 100

0.00018217 1.8037e-07

and still varies

burnin = 1600, T = 400, N = 100 (48 sec)

$\bar{g}^* = 0.00033$ $Var(\bar{g}^*_n) = 6.0104e-07$

varies still.

burnin = 0, T = 2000, N = 100 (150 sec)

$\bar{g}^* = 5.27773e-05$ $Var(\bar{g}^*_n) = 1.5801e-08$

$\bar{g}^* = 6.4184e-05$ $Var(\bar{g}^*_n) = 1.5136e-08$

2.6731e-05 N = 1 (2 sec)

9.6657e-05

1.511e-05

1.0546e-05

3 obs: 1) burnin doesn't matter (why? b/c doesn't impact FEV)

2) longer T decreases $\bar{g}^*$

3) decreases dispersion across n

} Ryan's conjecture was right.

Moving on to MSE:

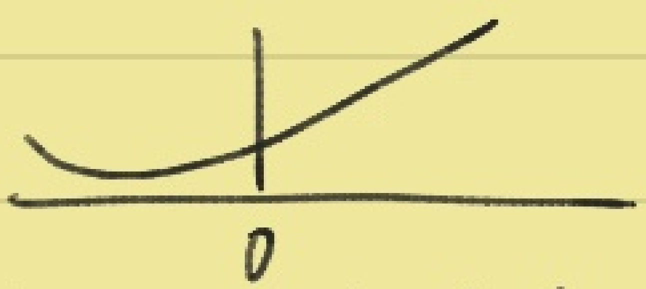
Note $MSE(y_j) = \frac{1}{T} \sum_{t=t_0}^{T} (y_{j,t} - \hat{y}_{j,t})^2$ (Brandt & Evans 2006)

$MSE = FEV$ b/c $FE = X_t - X_{t|t-1}$

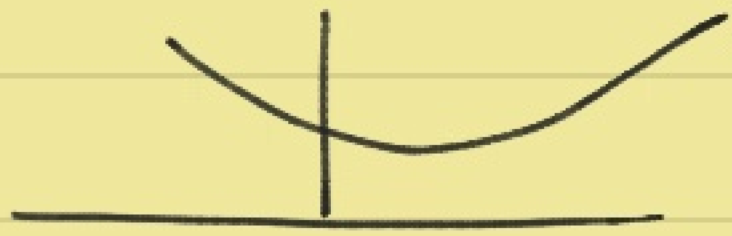
$Var(FE) = E[FE^2]$ = mean squared error

plotting this thing has a very clear shape:

it has a 2nd deriv > 0 (convex) and takes a min



either slightly to the left of 0



or slightly to the right

⇒ but maybe this means that if I increase T, the min will be pushed to 0 "from both sides"?

Any interesting finding: the more volatile the environment (the higher $\Sigma = VC(\text{shocks})$) the lower $\bar{g}^*$ (more negative)

Ok so what Ryan had had in mind was something like: sim Y using $\bar{g} = 0.145$.

Now, for different values of $\bar{g}$, compute forecasts of Y_{t+1} at each t, allowing only $\bar{\pi}$ to update and thus fcsts, but not Y. What's the $\bar{g}$ that mins the FEV?

→ For this I obtain $\bar{g}^* = 1.003e-05$ $Var(\bar{g}^*) = 3.357e-16$

whereas for my exercise I get $7.2424e-05$ $1.7218e-8$

(burnin $= 0$, $T = 2000$, $N = 100$)

The figs show that ideally you'd get $\approx -1.5 \cdot 10^{-3}$

→ and this is consistent no matter how long T is

(actually that's not too surprising given that here T doesn't make the FEV larger b/c the data is created given $\bar{g} = 0.145$)

Why am I getting negative values in both exercises?

Why is a $\bar{g}^* = 0$ not good? What other notion to use?

Given the relationship between the gain and the Kalman gain, is there no KF-related notion of optimality we could use?

→ In fact, for Ryan's minFEV, I get $\bar{g}^* < 0$ even when the DGP is RE!

[Scheinkman & Xiong, 2003 "Overconfidence & Speculative Bubbles"
setup on optimism & pessimism on financial markets
→ This stuff relates to my confirmation bias idea

<u>Understand why minFEV</u> <u>23 Nov 2015</u>

<u>Ryan's method</u>: y = generated by gain = 0.145

Now set $k = \bar{g}^* = \text{argmin FEV}(y(\bar{g} = 0.145))$.

<u>My method</u>: $k = \bar{g} = \text{argmin FEV}(y(\bar{g}))$

1.) I don't understand why Ryan said that I didn't go all the way to the fixed point. To me it seems like my method does go to the fixed point, his doesn't!

2.) Why did Ryan call his method "analog to RE"?

Interpretation of Ryan's method:

"Data is what it is. Let me update my beliefs so that they have as little errors given the structure of the data as possible."

→ I think that's why: b/c I wanna choose the optimal gain such that my expectations are close to model-consistent.

Interpretation of my method:

"Give me the gain that makes my beliefs model-consistent, internalizing that my beliefs affect the DGP."

→ to me my way still seems closer to a fixed point.

What would I expect the two methods to yield?

- Ryan expected his method to yield a lower $\bar{g}^*$ than $\bar{g} = 0.145$, the one using which data was generated. But why should it? This means Ryan expected that FEs are smaller when $\bar{g} < \bar{g}^{DGP}$. I don't think this is a general statement, i.e. I think he only expected that compared to $\bar{g}^{CEMP} = 0.145$. Why do I think that?

B/c fcsts are $\hat{E}_t \pi_{t+1} = \bar{\pi}_{t-1} + b_1 s_t$

where $\bar{\pi}_t = \bar{\pi}_{t-1} + k_t^{-1}(\pi_t - (\bar{\pi}_{t-1} + b_1 s_{t-1}))$

- So the statement is that even as data g is generated w/ $k_t = \bar{g}^{CEMP}$, that leads to FEs b/c the PLM doesn't coincide w/ the ALM. So if $k_t = \bar{g}^{CEMP}$ and FEs switch sign, that means that by lowering $\bar{g}$, we could lower FE b/c agents are overupdating.
- On the other hand, even $k_t = 0$ will not be optimal, b/c then $\bar{\pi}_t = \bar{\pi}_{t-1} = \bar{\pi}$, so $\hat{E}_t \pi_{t+1} = \bar{\pi} + b_1 s_t$ which yields permanent FEs when the data is generated by a gain of $k = \bar{g}^{CEMP}$. If $k_t = 0$, and initialized at RE, $\hat{E}_t \pi_{t+1} = 0 + b_1 s_t = $ RE fcst. So if the DGP is RE, then $\bar{g}^*$ better be 0, and it's a problem that that's not what I find!
- Also, this shows that for the learning DGP, $k^* = 0$

cannot be optimal.

So for Ryan's method, we'd expect an interior k^*.

What does it mean if this is < 0?

$$\text{fcsts are } \hat{E}_t \pi_{t+1} = \bar{\pi}_{t-1} + b_1 s_t$$

$$\text{where } \bar{\pi}_t = \bar{\pi}_{t-1} + k_t^{-1}(\pi_t - (\bar{\pi}_{t-1} + b_1 s_{t-1}))$$

If $k<0$, then if I underestimated π_t (fcst $< \pi_t$)

then I lower $\bar{\pi}$, that is, I lower my fcst even further. So that makes no sense!

$\Rightarrow$ it must be that these negative values are a result of a coding error, which potentially also explains why RE DGP doesn't give me $k^* = 0$.

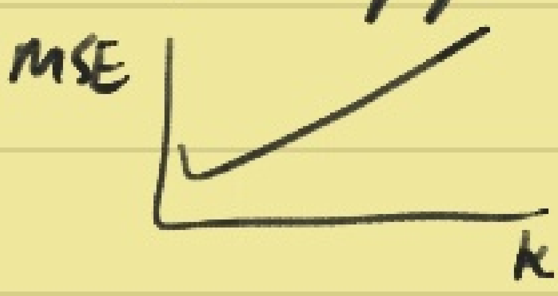
- So what do I expect my method to give?

Reinterpretation of my method: "give me the gain that gives me the sequence of Y for which the FEV given that gain is minimized" $\rightarrow$ i.e. "find the gain $\bar{g}^*$ that generates the sequence Y w/ minimal FEV associated w/ it"

→ but this should give you $k^* = 0$ b/c then the $\bar{\pi}_0 = 0$ means that PLM = ALM = RE, which again shows you why this isn't a good notion of an optimal gain!

⇒ Ryan's notion is better than mine

→ in fact I think Branch & Evans are doing something similar when estimating the gain, so I should obtain figures that look like theirs:



But: neither my nor Ryan's should give $k^* < 0$.
So there's sthg wrong in both!

So now I'm focusing on Ryan's method alone:

1) I'm getting $\approx 70\%$ of $\bar{g}^* < 0$ (!)
2) Increasing T makes 1) $\bar{g}^* \to 0$ from below 2) shrinks $\mathrm{Var}(\bar{g}^*_n)$
3) Increasing N doesn't do anything

→ in any case, $\bar{g}^* < 0$ about 60-70% of the time!

More observations:

1) If I make the shocks be meaned around ± some number, k^*s become positive!

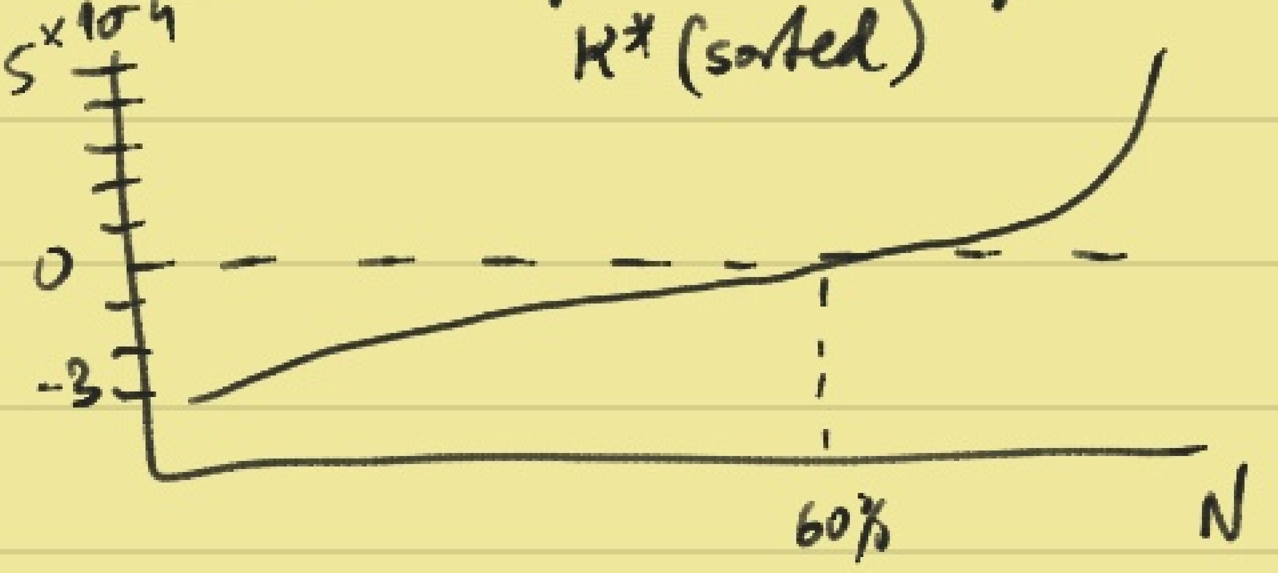
2) Changing Σ doesn't really do anything.

Why is this happening? And why doesn't the sign(mean(innovs)) matter? (→ i.e. why doesn't it matter whether shocks are centered around 1 or −1?)

→ Try shutting off all shocks except 1 → see how that affects k^*.

If in the objective function, I shut the effect of shocks on FE off, I obtain all $k^* > 0$

If you $T\uparrow$, $k^* \to 0$ but always from below and the distrib of k^* always is skewed like this:



Reading Gerko 2019 — 2 Dec 2019

I just understood something: $T(\phi)$ is the mapping from the PLM to the ALM. That is, given a belief ϕ, what ALM does the PLM imply, i.e. how would the model expect observables to evolve (absent shocks) given those beliefs?

$T(\phi) = E(ALM)$ (it just disregards shocks, answering the question: "w/ these beliefs, on average, what will observables do?")

The E-stability condition is the differential equation

$$\frac{d}{d\tau}(\phi) = T(\phi) - \phi \qquad (2.8, \text{EH p. } 31)$$

$E(ALM(\phi)) - PLM$ (either in both cases w/, or in both cases w/o observables)

CEMP's criterion $\theta_t = PLM - E(ALM)$, so it's really the E-stability diff. equation $\theta_t = \frac{d}{d\tau}(\phi)$ (!)

Susanto meeting 2 Dec 2019

Work after 3 Dec 2019

Ball 1994 AER

Deflation causes recessions as we thought.

But disinflation (changing the grth of money ↓)

causes a boom.

I think the idea is this:

$$x_t = m_t - E_t m_{t+1} \leftarrow \text{all future } m^s$$ (roughly)

↓ deflation causes x_t ↓

but disinflation is $\dot{m}$ ↓ so m_t doesn't change,

only m_{t+1} does (here's where credibility comes in: $E_t m_{t+1} = m_{t+1}$)

↳ this is exactly the contractionary $E_t\pi_{t+1}$ I get!

$$x_t = -\sigma \phi_\pi \pi_t + \sum_{T=t}^{\infty} \beta^{T-t} \left((1-\beta) x_{T+1} + \underbrace{\sigma(1-\beta\phi_\pi)}_{<0} \pi_{T+1} \right)$$

I think that the intuition is that a

credible disinflation means a movement in expectations

only, but no current values.
I think Whelan's intuition is that when $E_t \pi_{t+1} \downarrow$,
$P_t \downarrow$ a bit and so consumers are richer ($\frac{M_t}{P_t} \uparrow$)
$\rightarrow x_t \uparrow$

And this is also what Ball's intuitive proof circles around: firms choose lower prices <u>b/c they anticipate</u> future decreases in the increase of m.

↳ So, like for me, it's all about expectations moving and credibility: ppl have to believe mon. pol. (in my case, know & believe the Taylor-rule).

One could also summarize Ball's argument as
A disinflation has two effects: 1) contractionary via the interest rate 2) expansionary via expectations
For the expectations channel to dominate, Ball figured you need to do policy quickly, so "1) doesn't get to move".

In my case, you just need expectations to move strongly enough.

Whelan adds 2 things:

- Since Ball himself said that disinflationary booms don't fit the data well, it must be that CB's have credibility problems.
  - → But, though likely during Great Inflation, is it still likely now?
- Something else must be wrong w/ the NK model.

One thing I'm a little worried about is that E(·) moving strengthens the Ball effect, and E(·) move a lot when unanchored (again) which is exactly a measure of the CB not being credible. — ah but that may be fine actually, b/c I think Ball needs credibility for E(·) to move. I can get them to move otherwise.

Instrument instability seems to happen <u>4 Dec 2019</u>

here in the sense that expectations are the instrument.

In particular, the FE oscillates.

Let's try to write out the fcst error. It's

$$FE_{t-1} = \pi_t - (\bar{\pi}_{t-1} + b_1 s_{t-1})$$

where $\bar{\pi}_{t-1} = \bar{\pi}_{t-2} + k^{-1}(\underbrace{\pi_{t-1} - (\bar{\pi}_{t-2} + b_1 s_{t-2})}_{FE_{t-2}})$

$$FE_{t-1} = \pi_t - \left[\bar{\pi}_{t-2} + k^{-1} FE_{t-2} + b_1 s_{t-1}\right]$$

$$FE_{t-1} = \pi_t - \left[\bar{\pi}_{t-3} + k^{-1} FE_{t-3} + k^{-1} FE_{t-2} + b_1 s_{t-1}\right]$$

$$FE_{t-1} = \pi_t - b_1 s_{t-1} - \underset{\overset{\shortparallel}{0}}{\bar{\pi}_0} - k^{-1} \sum_{s=0}^{t-2} FE_s \quad \Big| \pm FE_j$$

$j = t-2, \dots, "1"$

$$\Delta FE_{t-1} = \underbrace{\pi_t - b_1 s_{t-1}}_{} - k^{-1} \sum_{s=0}^{t-2} \Delta FE_s$$

ignore these and switch $t-1$ to t for simplicity

$$\Delta FE_t = -k^{-1}\left[\Delta FE_{t-1} + \Delta FE_{t-2} + \dots\right]$$

so the weights are k^{-1} for all and 1 for ΔFE_t

or $k < -1$ for ΔFE_t and 1 for all the rest.

$k \Delta FE_t + \Delta FE_{t-1} + \Delta FE_{t-2} + \dots$ leads to the characteristic equation

$$k x^n + x^{n-1} + \dots + x + 1 = 0$$

If we only had one lag, this would be

$$k x + 1 = 0$$

$$x = -k^{-1} > -1$$ That would be stable. but oscillating

Two lags

$$k x^2 + x + 1 = 0$$ $$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$x_{1,2} = \frac{-1 \pm \sqrt{1-4k}}{2k}$$

For the roots to be real $1 - 4k \overset{!}{>} 0 \Rightarrow 1 > 4k$

$k \overset{!}{<} \frac{1}{4} = 0.25$ It usually is, even CEMP's.

If $k = 0.25$, $x_1 = x_2 = -\frac{1}{2k} = -2$

→ that would be unstable.

For $k \in (0, 0.25)$, both roots are < -1 always!

So the system is unstable.

My guess is that the more lags you include, the closer you will get to stability, <u>but</u> you will have (potentially all) roots <0, which is why we get the oscillation.

But what this doesn't account for is the role of ψ_π in getting the oscillation. What seems to be clear is that the oscillations in FE are driving it. So in a sense I'm not even sure if it's instrument instability or simply instability.

The connection to ψ_π must come via the role of i in π via x.

$$\left.\begin{aligned} x &= -\sigma i_t + E(\text{stuff}) \\ \pi_t &= \kappa x_t + E(\text{stuff}) \\ i_t &= \psi_\pi \pi_t \end{aligned}\right\}$$

In period 1, a shock hits, E moves a little, (x, π) move, i moves.

Evening of period 1: E(·) adjusts. 2 things influence the adjustment: 1) $\kappa = \bar{g}$ the size 2) ψ_π the size of Ball's disinflationary boom effect (<u>direction</u> of E-adjustment)

$FE = \pi_t - (\bar{\pi}_{t-1} + b_1 s_{t-1})$

↑k governs the change here

ψ_π governs the change in <u>both</u>, and the bigger it is, the more E( ) move compared to π, opening up FEs.

→ it's as if E(·) were the policy instrument?

Changing i works partly via its effect on x_t and π_t today, but its main effect is via E(·).

So in that sense it's like instrument instability b/c the instrument becomes unstable, <u>but</u>: it's also not like instr. instab. b/c a too high ψ_π makes FEs unstable thereby rendering the objective variables unstable too.

Ryan meeting 4 Dec 2019

- Gerko: send to Ryan

- Data-IRFs oscillated for Ryan early on → so it's not quite the case that empirical IRFs never oscillate ...

- L'Huillier estimated $\hat{\psi}_\pi \approx 1.001$ or sthg

  Note: it's Blanchard, L'Huillier Lorenzoni News & Noise, AER 2013 and they get $\hat{\psi}_\pi = 1.0137$, $\hat{\psi}_x = 0.005$

- Try $\psi_\pi < 1$ → it seems like if for a fwd-looking system $\psi_\pi > 1$ gives stability, for a bw-looking one, $\psi_\pi < 1$ will.

- Consider flex price model

$$\left.\begin{aligned} r_t &= i_t - E_t[\pi_{t+1}] \\ i_t &= \phi \pi_t \end{aligned}\right\} \text{ do expectations pan out similarly here?}$$

- The big picture question is:

  Do we take the model's implications seriously and explore what they imply for policy?

  OR: do we change sthg about the model? 2 options

  1) change E-formation 2) change policy (e.g. have $E(\pi)$ in TR instead of π)

If you change policy, then you can make a statement like: "central bankers say they have $E(\pi)$ in TR, and look, indeed it works better than π"

If you change $E(\cdot)$-formation, you can say: "std learning implies this, but here's a learning that works".

Think of ways we can change policy

· $E(\pi)$ instead of π in TR

Think of ways we can change E-formation.

<u>Work after</u> <u>5 Dec 2019</u>

One thing I've tried is to add interest rate smoothing, $\rho>0$. It doesn't change much b/c on the net it acts like ρ_i, the persistence of the mon pol shocks, except that it has a contractionary effect (stuck w/ initial shock for longer) and an expansionary effect (stuck w/ expansionary policy reaction for longer). In gen, these two seem to balance so we still have the overshooting.

Peter meeting 5 Dec 2019

A big-picture comment:
What you might worry about in learning:
Systematic pattern of FEs
→ serially corr errors in the same direction

Peter Ireland similar paper to Ball:
Stopping Inflations, Big & Small 1997 JMCB

Ric. Eq. seems to hold today? Chris Sims
30-yr T-bill i-rate is low despite high debt

→ ask if
alternative expectation-formation schemes that
deliver more appealing dynamics for
E(·) and thus for observables.

Work after 5 Dec 2010

Reading Townsend (1983): "Forecasting the forecasts of others" which seems to be the first paper to introduce dispersed info!

One claim he makes is that w/ a normal signal extraction problem:

- forecasts $E[v_t | \Omega_t]$ (info set) is correlated w/ $E[v_{t-1} | \Omega_{t-1}]$ (i.e. fcsts serially corr) b/c both contain v_{t-1}
- FEs $[E(\theta_t | \Omega_t) - \theta_t]$ will be serially corr too b/c it'll be an MA of past random vars v_t and ε_t.

Another is that in a simple model of fcsting the fcsts of others, FEs of those forecasting the fcsts of others (market 2) exhibit <u>damped oscillations</u> (!)

$\Rightarrow$ this may be part of the reason for me too b/c in this learning model agents don't know other people's fcsts but are implicitly fcsting them when fcsting aggregates.

To do's:

(1) Add $E(\pi)$ instead of π in TR

(2) Try $\psi_\pi < 1$

(3) The Townsend analysis suggests that IRFs to EE-learning won't give oscillations b/c there agents aren't testing each others' tests

(4) Do IRFs to vector learning from the spinoff LH paper

(5) Figure out alternative learning that exhibits no oscillations

- Can Bayesian learning do it?

↳ refs in Ryan's class reading list

Susanto commented on oscillating IRFs: 6 Dec 2019

1) They can and do arise more frequently than you'd think b/c published IRFs look "nice" b/c a lot of work goes into making them nice

2) check local projections which estimates the MA process directly. Since a VAR estimates an AR and imposes

restrictions on the VC matrix (and the AR-matrix), it's likely to give you more smooth IRFs, while those of local projections are likely to be more choppy. Check Valerie Ramey's Handbook chapter on IRFs! It also has a bunch of replication files.

Adding $E(\pi)$ to TR <u>7 Dec 2019</u>

$$x_t = -\sigma\left(\psi_\pi \hat{E}_t \pi_{t+1} + \psi_x x_t + \underbrace{\rho i_{t-1} + \bar{i}_t}_{} \right) + \Sigma(\cdot)$$

clearly the shocks will be the same, ignore 'em

$$(1+\sigma\psi_x)x_t = -\sigma\psi_\pi \hat{E}_t \pi_{t+1} + \Sigma$$

$$(1+\sigma\psi_x)x_t = -\sigma\psi_\pi \hat{E}_t \pi_{t+1} + \hat{E}_t \sum_{T=t}^{\infty} \beta^{T-t}\Big[(1-\beta)x_{T+1} + \sigma\pi_{T+1} - \sigma\beta\big(\psi_\pi \hat{E}_{T+1}\pi_{T+2} + \psi_x x_{T+1} + \text{shocks}\big)\Big]$$

$$(1+\sigma\psi_x)x_t = -\sigma\psi_\pi \underline{\hat{E}_t \pi_{t+1}} + \hat{E}_t \sum_{T=t}^{\infty}\beta^{T-t}\Big[(1-\beta-\sigma\beta\psi_x)x_{T+1} + \sigma\pi_{T+1} - \sigma\beta\psi_\pi \underline{\hat{E}_{T+1}\pi_{T+2}}\Big]$$

now these two pose difficulties

b/c I don't know if I can use LIE on $\hat{E}_t$ or not.

Let's see if we can shove σi_t into the expectation instead.

$$x_t = -\sigma i_t + \hat{E}_t \sum_{T=t}^{\infty} \beta^{T-t} \left[(1-\beta) x_{T+1} + \sigma \pi_{T+1} \right]$$

$$+ \hat{E}_t \sum_{T=t}^{\infty} \beta^{T-t} \left[-\sigma \beta i_{T+1} \right]$$

$$= \hat{E}_t \sum_{T=t}^{\infty} \beta^{T-t} \left[-\sigma i_T \right]$$

$$\Rightarrow x_t = \hat{E}_t \sum_{T=t}^{\infty} \beta^{T-t} \left[(1-\beta) x_{T+1} + \sigma \pi_{T+1} - \sigma i_T \right]$$

$$= \hat{E}_t \sum_{T=t}^{\infty} \beta^{T-t} \left[(1-\beta) x_{T+1} + \sigma \pi_{T+1} - \sigma \left(\psi_\pi \hat{E}_T \pi_{T+1} + \psi_x x_T \right) + \text{shocks unchanged} \right]$$

$$(1+\sigma\psi_x) x_t = \hat{E}_t \sum_{T=t}^{\infty} \beta^{T-t} \left[(1-\beta) x_{T+1} + \underbrace{\sigma \pi_{T+1} - \sigma \psi_\pi \hat{E}_T \pi_{T+1}}_{\text{LIE is still a question}} + \text{shocks unchanged} \right]$$

but at least this expression is neater.

So $\hat{E}^i_t \hat{E}^i_{t+1} = \hat{E}^i_t$, i.e. the idio $\hat{E}^i$ fulfills LIE b/c that is anticipated utility. There are 2 questions: if LIE holds for $\hat{E}^i$, why don't HHs use the recursive representation?

→ Preston would argue that b/c then they would ignore their wealth b/c they don't know they're identical. ok fine so they don't use LIE and so we get the LH expression when we aggregate.

But it seems to me - again! - from Preston 2005 p 16 that

that $\hat{E}$ doesn't satisfy LIE – not on the aggregate.

So also CEMP define anticipated utility as $\hat{E}^{i}_{t-1} \bar{\pi}_T = \bar{\pi}_t$

i.e. as an idiosyncratic statement.

However it seems to me like maybe actually even the agg $\hat{E}(\cdot)$ satisfied it somehow: so when we compute expectations I think we make use of it?

In particular, the modeler knows, as CEMP state, that $\hat{E}^{i}_t = \hat{E}_t$, i.e. that agents hold identical beliefs. So then $\hat{E}_t$ will fulfill LIE too. So maybe it's really just the case that agg beliefs just "inherit" the discounted sums.

Anyway – the usual confusion about LH is there: it seems like everything fulfills LIE and yet still since LH forecasts are optimal for agents who don't know that they're identical, agg $\hat{E}$ "inherit" them.

Will do so after that detour:

$$(1+b\psi_x)x_t = \hat{E}_t \sum_{T=t}^{\infty} \beta^{T-t}\Big[(1-\beta)x_{T+1} + \underbrace{b\pi_{T+1} - b\psi_\pi \hat{E}_T \pi_{T+1}}_{=(b-b\psi_\pi)\pi_{T+1} \text{ here}} \text{ + shocks unchanged}\Big]$$

$$(1+b\psi_x)x_t = \hat{E}_t \sum_{T=t}^{\infty} \beta^{T-t}\Big[(1-\beta)x_{T+1} + b\underbrace{(1-\psi_\pi)}_{\text{well this will always be } <0}\pi_{T+1}\Big]$$

So then: $x_t = \underbrace{\frac{1}{1+b\psi_x}\left[b(1-\psi_\pi),\ 1-\beta,\ 0\right]}_{=g_{xb}} f_\beta$ $\qquad g_{xa} = [0\ 0\ 0]$

$$\pi_t = \kappa x_t + \left[(1-\alpha)\beta,\ \kappa\alpha\beta,\ 0\right] f_\alpha$$

$$\pi_t = \underbrace{\frac{\kappa}{1+b\psi_x}\left[b(1-\psi_\pi),\ 1-\beta,\ 0\right]}_{\kappa\, g_{xb}} f_\beta + \underbrace{\left[(1-\alpha)\beta,\ \kappa\alpha\beta,\ 0\right]}_{=g_{\pi a}} f_\alpha$$

$$x_t = g_{xb}\, f_\beta + \underbrace{g_{xa}}_{\to [0\ 0\ 0]} \cdot f_a$$

$$\pi_t = \kappa \cdot g_{xb} \cdot f_\beta + g_{\pi a} \cdot f_a$$

+ shocks that are unchanged

$$i_t = (\psi_\pi g_{\pi a} + \psi_x g_{xa}) f_a + (\psi_\pi g_{\pi b} + \psi_x g_{xb}) f_\beta$$

are learning the <u>distribution of shocks</u>, $g(\cdot)$.

For Koz et al, this is crucial b/c when a tail event occurs, it leads to agents to estim a shock distrib $\hat{g}(\cdot)$ w/ <u>fat tails</u> (potentially asymmetric?)

They estimate $\hat{g}(\cdot)$ using a <u>normal kernel density estimator</u>, I guess the most std. non-parametric estimator:

x_t? not clear but can check kernel estimators

$$\hat{g}_t(x) = \frac{1}{n_t} \sum_{s=0}^{n_t - 1} \Omega(x - x_{t-s}, \Sigma_t)$$

or mean, $\bar{x}$? (see Baum lecture 3)

\# obs at time t

multivariate normal density

VC matrix (aka "bandwidth matrix")

x = $d \times 1$ shock vector

Σ_t = diagonal matrix, I think like this:

$$\underset{d \times d}{\Sigma} = \begin{pmatrix} \hat{\sigma}_1 & & & \\ & \hat{\sigma}_2 & & \\ & & \ddots & \\ & & & \hat{\sigma}_d \end{pmatrix} \cdot \left(\frac{4}{(2+d)n_t} \right)^{\left(\frac{1}{4+d}\right)}$$

where $\hat{\sigma}_j$ = sample std dev of shock j

So 2 main differences to adaptive learning:

1) here agents learn about shocks, not observables

2) learning is nonparametric b/c they learn a distrib

↳ (1) is not so interesting for me b/c exog. shocks have no connection to central bank credibility, so that one can no longer talk about anchoring

↳ I expect (2) to be more interesting, yet maybe not too different to my setting (qualitatively) b/c

I can imagine a generalization of my adaptive learning framework in which agents don't know the distrib of the LOM (they don't know that gx is linear and normal) ⇒ that would be similar to learning slope and constant I guess except it would be less restrictive on their "prior" if you will.

→ But one could try it out!

What is not clear to me is what would be the analogue to the gain? The bandwidth matrix?

Read Collins-Dufresne et al (2016)

Honestly, Bayesian learning seems very analogous to adaptive (frequentist) learning except that the updating rule is Bayes' rule, the prior matters AND one difference that might matter is that while in adaptive learning, the anticipated utility ass means that in every period, agents treat the PLM as if it was the ALM, Bayesian learning can account for the uncertainty around the correctness of the PLM (this is referred to as parameter uncertainty) and it often refers to the case when agents first using an average of different models

→ can this mitigate overshooting?

Or is this simply the Bayesian counterpart to the gain / anchoring?

The problem is that this is an asset pricing paper and doesn't describe how Bayesian learning works, nor does it give references.

Evans, Honkapohja & Sargent (2013)

(In the book Macro at the Service of Public Policy – which I haven't downloaded but I did save the ref.)

Muth market model (6.1)

$$p_t = \mu + \alpha E^*_{t-1} p_t + \delta z_{t-1} + \eta_t \,; \quad \eta_t \sim WN(0, \bar{\sigma}^2_\eta)$$

$$z_t = \rho z_{t-1} + w_t \qquad w_t \sim iid(0, \sigma^2_w)$$

$$\mu = 0$$

REE model for (6.1) is

$$p_t = \bar{\beta} z_{t-1} + \eta_t \qquad w/ \ \bar{\beta} = \frac{\delta}{1-\alpha}$$

Now EHS consider a constant parameter learning rule that's estimated using Bayesian techniques:

the PLM is

$$\sigma^2_\eta \neq \bar{\sigma}^2_\eta$$

$$p_t = b_{t-1} z_{t-1} + \eta_t \qquad w/ \ \eta_t \sim N(0, \sigma^2_\eta)$$

Agents' prior distrib for β is $\beta \sim N(b_0, V_0)$ and updates are

$$b_t = b_{t-1} + \frac{V_{t-1} z_{t-1}}{\sigma^2_\eta + V_{t-1} z^2_{t-1}} (p_t - b_{t-1} z_{t-1})$$

$$V_t = V_{t-1} - \frac{z^2_{t-1} V^2_{t-1}}{\sigma^2_\eta + V_{t-1} z^2_{t-1}} \quad \text{using the Kalman filter.}$$

Prop. 6.1 Bayesian analog to E-stability

Convergence to REE w/ prob 1 if $\alpha < 1$

and $V_t = \dfrac{\sigma_\eta^2}{(t+1)S_t - z_t^2} \rightarrow 0$ irrespective of whether σ_η^2 is correct!

$\downarrow$

WTF is S_t...?

I'm a little suspicious of this b/c it says that if $V_t \rightarrow 0$ then the Kalman gain goes to zero too and you don't get that w/ the KF (instead $\Sigma_{t|t} \rightarrow \Sigma$, the st. st. and so the gain $K_{t|t}$ goes to K, the st. st gain) $\Rightarrow$ this is normally how the KF is analogous w/ constant gain learning and it always keeps a tiny variance around RE.

But maybe they are looking at dgain learning analogues here – that might be what they mean w/ "constant param learning". In the following section, they consider time-varying param β.

It is useful to understand how time-varying param (TVP) models are related to misspecification of fcst rules. When agents think that βs may be time-varying they entertain the idea that they may be using a misspecified model. This is intimately connected to cgain learning b/c agents prefer to track the params because they are never sure whether a structural change may render their converged fcst-ing rule misspecified.

EHS show w/ a simple example (where agents aren't sure whether a constant param PLM or a TVP PLM is the right one) that if feedback effects from beliefs are sufficiently strong then agents may select the non-RE model b/c E(·) induce actual drifts in the ALM that induce the MSE of the non-RE PLM to be lower than the RE one. (This is like a self-confirming learning result.)

A note: EHS claim on p. 105 that

they guess that for this TVP learning case the stable param space will be $-1 < \alpha < 0.5$ b/c under LS learning:

1) $0.5 < \alpha < 1$ may yield slow convergence to RE

2) "$\alpha < -1$ a possible problem of overshooting can emerge when agents overparametrize the PLM"

$\Rightarrow$ oh really??

So when the feedback from expectations is negative and sufficiently strong, least-squares learning may yield overshooting???

You don't say!