

Work after

9 Nov 2019

Form the FEs correctly:

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

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

$$\left. \begin{array}{l} FE_t^m = \pi_{t+1} - f_t^m \\ FE_t^e = \pi_{t+1} - f_t^e \end{array} \right\} \text{so in either case I just need to subtract } F_t \text{ from } y_{sim,t+1} \text{ and note that these } F_t \text{ are realized at } t+1.$$

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 fest of π .

The problem is that the FE coming out of sim-learn

- 1) always changes, despite IRF-ing & averaging
- 2) there are diffs b/wn FE_{shocked} & FE_{unshocked} even before I impose the thresov δ (!)

FEs are solved.

14 Nov 2015

I think that the cross-coupling of fots is well-understood: when (gain), you update your foot too much and so your FE switches sign and oscillates.

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 /

crosscoupling w/ a sufficiently low gain

i.e., 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 b/wn $f_a \& f_b$ & overshooting too!

$$f_A = \frac{a}{1-\alpha\beta} + b (I_{nx} - \alpha\beta h x)^{-1} s$$

What is $\frac{1}{1-\alpha\beta} = 50.2573$ and $\frac{1}{1-\beta} = 100$

for $\alpha = \beta = 0.95$?

$$(F_A - \alpha\beta h x)^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2.7275 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0.0049 & 2.5267 & 1.9678 & 1 \end{bmatrix}$$

$$\text{and } (F_B - \beta h x)^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2.7639 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0.0050 & 2.5643 & 1.9776 & 1 \end{bmatrix}$$

ha! The diff in f_A & f_B is most pronounced in the part that comes from the intercept!

I think M_N would change a bit as α moves away from 1 (M_0)

$$\alpha = 0.5, \beta = 0.95$$

$$\frac{1}{1-\alpha\beta} = 1.0802 \quad \text{and} \quad \frac{1}{1-\beta} = 100$$

$$(F_4 - \alpha \beta h x)^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1.4225 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0.0015 & 0.6817 & 0.7388 & 1 \end{bmatrix}$$

$$\text{and } (F_4 - \beta h x)^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2.7639 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0.0050 & 2.5643 & 1.9776 & 1 \end{bmatrix}$$

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

\Rightarrow the lower α (the higher k , the less price children)
the more b_0 loads on the intercept both in absolute terms
(it reacts more) and relative terms (vs. the slope)

\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_+ up for gains.

Let's interpret

$$\begin{matrix} \alpha\beta \cdot h_x & \text{vs} & \beta \cdot h_x \\ \left[\begin{array}{cccc} 0 & 0 & 0 & 0 \\ 0 & 0.297 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0.0025 & 0.48 & 0.7388 & 0 \end{array} \right] & \begin{matrix} r^n \\ \bar{r} \\ u \\ i_{t-1} \end{matrix} & \left[\begin{array}{cccc} 0 & 0 & 0 & 0 \\ 0 & 0.554 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0.005 & 0.56 & 1.4776 & 0 \end{array} \right] \end{matrix}$$

2 differences

- 1) effect of shocks on i_t
 - 2) effect of \bar{r} 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 $t=2$

$$A_a(3,1) = 0.5928 \rightarrow i \uparrow \text{if } f_a(1) \uparrow$$

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

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

But why?

if $\gamma_x \uparrow$ (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 $\gamma_x = 5$.

When $\gamma_x = 0$, 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

$$A_a(3,1) = \gamma_{\bar{a}} A_a(1,1) + \gamma_x A_a(2,1)$$

$$A_b(3,1) = \gamma_{\bar{b}} A_b(1,1) \quad "0"$$

these are true!

⇒ 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, i has depressed a lot.

At $t=2$, since $\pi \uparrow$ (but is still < 0), i is depressed below $0.6 \cdot 1^{\uparrow \delta}$ (i^{\uparrow} 's only ≈ 0.1) but it's not depressed as much

Puzzling i-response

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

$\downarrow \quad \uparrow$

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 what's going on is that $E[\bar{\pi}]$ are pushing stuff up but i^* is pushing them down, and i reacts faster

Check: if $\bar{\pi}$ -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 $\pi_t \neq \bar{\pi}$ reaction to expectations, when RE doesn't have bias?

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

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

$$i_t = \gamma_\pi \pi_t$$

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

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

$$x_t = -\beta \gamma_\pi \pi_t + E_t x_{t+1} + \beta E_t \pi_{t+1}$$

$$x_t = -2\gamma_\pi [K x_t - \beta E_t \pi_{t+1}] + E_t x_{t+1} + \beta E_t \pi_{t+1}$$

$$X_t = -\gamma \pi_t [kx_t + \beta E_t \pi_{t+1}] + E_t x_{t+1} + \beta E_t \pi_{t+1}$$

$$(1 + \gamma \pi_t k) X_t = -\gamma \pi_t \beta E_t \pi_{t+1} + E_t x_{t+1} + \beta E_t \pi_{t+1}$$

$$X_t = \underbrace{\frac{1}{w} \beta (1 - \gamma \pi_t \beta) E_t \pi_{t+1}}_{< 0 (!)} + \frac{1}{w} E_t x_{t+1}$$

$\Rightarrow RE$ has it too, only

$$\Rightarrow \pi_t = \frac{k}{w} \beta (1 - \gamma \pi_t \beta) E_t \pi_{t+1} + \frac{k}{w} E_t x_{t+1} + \beta E_t \pi_{t+1}$$

[E]

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

more
as much

(for current params = 0.9298)

RE :

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

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

Learning

$$X_t = \ominus E_t \pi_{t+1}^{fa} + \ominus E_t \pi_{t+1}^{fb} + \circlearrowleft \ominus E_t x_{t+1}^{fa} + \oplus E_t x_{t+1}^{fb}$$

$$\pi_t = \oplus E_t \pi_{t+1}^{fa} + \circlearrowleft \ominus E_t \pi_{t+1}^{fb} + \oplus E_t x_{t+1}^{fa} + \ominus E_t x_{t+1}^{fb}$$

$$X_t = \ominus E \pi + \circlearrowleft \ominus E X$$

① *This is not a mistake*

$$\pi_t = \oplus E \pi + \oplus E X$$

②

$$\textcircled{1} \quad -\frac{\beta \gamma_{\pi} k \alpha \beta}{w} + \frac{1}{w} (1-\beta) = \frac{1-\beta - \beta \gamma_{\pi} k \alpha \beta}{w} < 0$$

$$\textcircled{2} \quad \left(1 - \frac{k \beta \gamma_{\pi}}{w}\right)(1-\alpha)\beta + \frac{k}{w} \beta (1-\beta \gamma_{\pi}) > 0$$

\Rightarrow why do we have this diff b/w RE & Learn? 15 Nov 2017

$$\beta \gamma_{\pi} k \alpha \beta < 1 - \beta$$

$$\beta \gamma_{\pi} k \alpha \beta + \beta < 1$$

$$(\beta \gamma_{\pi} k \alpha + 1) \beta < 1 \quad \text{but it's } 1.1150$$

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.

$$\text{RE: } \pi \text{ in } E(\pi): \quad \frac{k \beta + \beta}{\beta \gamma_{\pi}} = 0.7722 \text{ under current params}$$

$$\text{Learn: } \left(1 - \frac{k \beta \gamma_{\pi}}{w}\right)(1-\alpha)\beta + \frac{k}{w} \beta (1-\beta \gamma_{\pi}) = 0.33 \quad -11-$$

\rightarrow So yes, this is true

In fact, you can reason that in RE,

$\pi = E(\pi)$ only b/c only via
④

while in learning

$\pi = E(\pi)$ but part of π_{RL} is
④ → from f_1
⑤ → from f_2

Ryan meeting

15 Nov 2019

fix point where find the gain stat with $\text{Var}(\text{FE})$
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 !

Analogy to RE for the gain problem.

→ Pooya Molavi's JMP does RL, 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-pred

$$PLM = \bar{\pi}_{t+1} \text{ and that's it.}$$

0.2. My learning the slope

\Rightarrow in those contexts, do I continue to get the hiccups?

2. Didn't quite get to the bottom of RE is.

learning loading on EC.)

\rightarrow connect to those equations

3. These features can become worse if ΓY_π

$$\uparrow Y_\pi$$

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 $Y_\pi \uparrow$.

Here's how it works.

Work after

- 15 Nov: did "only-mean" Pm and "slope & constant"
- ✓ check that the latter is correct
 - ✓ print figs w/o cutting them off
 - polish explanation of $E(\cdot) \rightarrow z$
 - do the fixed point thing, use Molari

Reading Molari JMP

16 Nov 2019

I'm not superimposed 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, Molari calls this a pure CREE)
He shows (Thm 2 & 3) that Bayesian & adaptive learning coincide w/ a CREE in the LH
 \rightarrow of course b/c the CREE must be = REE if

the REE $\in \Theta$! This is why Molen says that the "id 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 Molen's concept of $\text{REE} \notin \Theta$, 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 \leftarrow I suppose time.

$$\text{Var}(x) = E(x^2) - E(x)^2$$

$$\text{If } x = \text{FE}, \text{ then } E(x) = 0, \text{ so } \text{FEV} = E(\text{FE}^2)$$

$$\text{Var}(x) = E \left[\underbrace{(x - E[x])^2}_{\text{FE}} \right] \quad \text{this is why}$$

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

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

\uparrow
fcast

Ok let's clarify one thing:

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

are the same thing?

Leaving that aside for a moment

- My posts are given by the $\text{PLM}(\bar{g})$
 - $\text{FE}_{t+1} = \pi_t - \text{PLM}_{t-1}^{e_t}(\bar{g})$
 - $\text{FEV}_{t-1} = E \left[(\pi_t - \text{PLM}_{t-1}^{e_t}(\bar{g}))^2 \right]$
- Let's take a general case: $\text{PLM}_{t-1}^{e_t}(\bar{g}) = \phi_{t-1} \begin{bmatrix} 1 \\ s_{t-1} \end{bmatrix}$
- $$\Rightarrow \text{FEV}_{t-1} = E \left[(\pi_t - \phi_{t-1}(\bar{g}) \begin{bmatrix} 1 \\ s_{t-1} \end{bmatrix})^2 \right]$$
- $$\phi_{t-1}(\bar{g}) = (\phi_{t-2} + \bar{g} (\pi_t - \phi_{t-2} \begin{bmatrix} 1 \\ s_{t-2} \end{bmatrix}))'$$

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 FER across time \rightarrow I make it
min FER for each history n

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

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

\Rightarrow Think more on this tomorrow!

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

17 Nov 2019

RE:

$$X_t = -\beta \gamma_{\pi} (K X_t + \beta E_t \pi_{t+1}) + E_t x_{t+1} + \beta E_t \pi_{t+1}$$

$$= -\beta \gamma_{\pi} K X_t - \beta \beta \gamma_{\pi} E_t \pi_{t+1} + E_t X_{t+1} + \beta E_t \pi_{t+1}$$

$$(1 + \beta \gamma_{\pi} K) X_t = \beta (1 - \beta \gamma_{\pi}) E_t \pi_{t+1} + E_t X_{t+1}$$

$$X_t = \frac{\beta (1 - \beta \gamma_{\pi}) E_t \pi_{t+1} + \frac{1}{1 + \beta \gamma_{\pi} K} E_t X_{t+1}}{1 + \beta \gamma_{\pi} K}$$

$$\pi_t = \frac{K \beta (1 - \beta \gamma_{\pi}) E_t \pi_{t+1}}{1 + \beta \gamma_{\pi} K} + \frac{K}{1 + \beta \gamma_{\pi} K} E_t X_{t+1} + \beta E_t \pi_{t+1}$$

$$\pi_t = \left(\frac{K \beta (1 - \beta \gamma_{\pi})}{1 + \beta \gamma_{\pi} K} + \beta \right) E_t \pi_{t+1} + \frac{K}{1 + \beta \gamma_{\pi} K} E_t X_{t+1}$$

Learning

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

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

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

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

matlab10 \rightarrow parameter values:

$$\cdot \frac{k\beta(1-\beta\gamma_\pi)}{1+\beta\gamma_\pi K} + \beta = \frac{k\beta - k\beta\beta\gamma_\pi + \beta + \beta^2\gamma_\pi K}{1+\beta\gamma_\pi K} > 0$$

$$\cdot -\frac{\beta\gamma_\pi K\alpha\beta}{1+\beta\gamma_\pi K} + \frac{1-\beta}{1+\beta\gamma_\pi K} \propto 1-\beta - \beta\gamma_\pi K\alpha\beta$$

For this to be positive, we need $\beta + \beta\gamma_\pi K\alpha\beta < 1$

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

$$\cdot \left(1 - \frac{k\beta\gamma_\pi}{1+k\beta\gamma_\pi}\right)(1-\alpha)\beta + \frac{k\beta(1-\beta\gamma_\pi)}{1+\beta\gamma_\pi K}$$

$$\propto (1+k\beta\gamma_\pi - k\beta\gamma_\pi)(1-\alpha)\beta + k\beta - k\beta\beta\gamma_\pi$$

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

$$\underbrace{\beta + \alpha\beta}_{\approx 1.5} + \underbrace{k\beta}_{\approx 20} - \underbrace{k\beta\beta\gamma_\pi}_{\approx \gamma_\pi \text{ even if } \gamma_\pi = 5} > 0$$

Ok, so now explain why, if I recursively substitute 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 $\hat{E}_t^i \hat{E}_{t+1}^i = \hat{E}_t^i$ (in fact, this is anticipated utility!) but not for the average expectation: $\hat{E}_t^i \hat{E}_{t+1}^i \neq \hat{E}_t^i \Rightarrow$ it's a little bit like the distinction b/w PLM & ALM b/c firms act based on $\hat{E}_t^i \hat{E}_{t+1}^i = \hat{E}_t^i$, i.e. knowing that LIE holds, but in the actual law of motion $\hat{E}_t^i \hat{E}_{t+1}^i$ turns out not to equal \hat{E}_t^i since updating happens!

19 Nov 2015

For Susanto, use

- from materials 10: "A more concise rephrasing"
- I think I wanna show IRFs from Dgair & again against RE for std params for the 3 shorts (take iid shorts & except mnpd.)

Ryan meeting

(500 years) 20 Nov 2015

↳ do for FER-min $T = 5 \cdot 400$
→ question: maybe this isn't ergodic $\Rightarrow \bar{g}^* \text{ is too large}$
Fabio Milani } have estimated gains
& Preston }

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

Ryan conjecture:

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

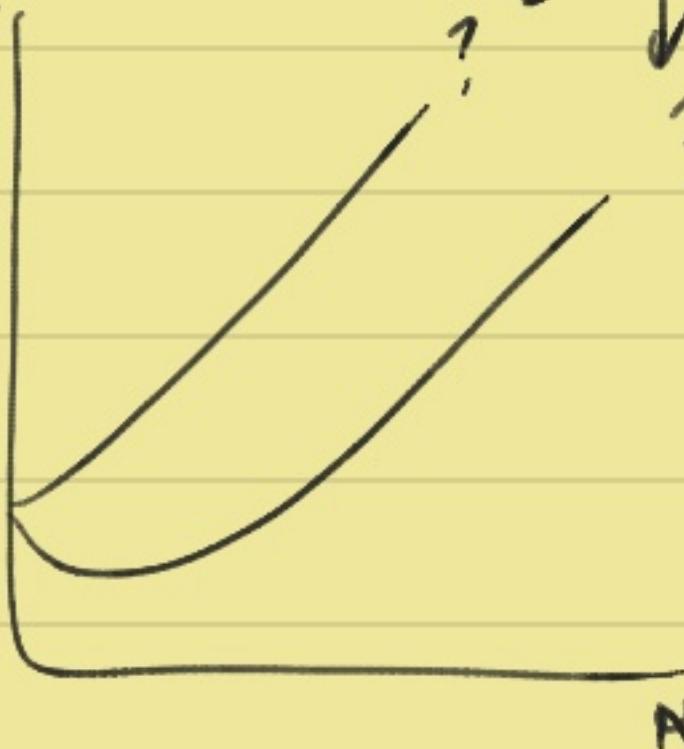
↳ But if this exercise gives $\bar{g}^* = 0$ then it's not the right model-based notion of what the \bar{g} should be)

→ 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

MSE



We don't know yet what π_{M3} fit 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!

→ Move slightly neg. α to see if it comes

back up. That is, if the MSE function looks like
this:



If not unique for reasonable K ,
can $\pi - E(\cdot)$ look absurd?

↳ 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 & - either from a model & optimality perspective or from the data. And the question is: supp. we have a reasonable κ , κ^R .

1) Do we get overshooting for that?

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

If overshooting is endemic for κ^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 $\beta = 0.95$,

Eusepi, Gramm, Preston (2019)
Limits

$g < 0.02$.

$\hat{g} = 0.05$

0.145

CAMP

0.0183

Milan (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 CAMP & avg estimate w/o CAMP = 0.04

Let's interpret the gain number:

- Euzgi & 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"

[Euzgi & Preston (2011) \rightarrow 500 quarters ($\Rightarrow 125$ years!)
avg w/o COMP \rightarrow 24 quarters ($\Rightarrow 6$ years)
COMP \rightarrow 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)

$$\hookrightarrow \min(\bar{g}) = 0.005$$

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

$$\hookrightarrow \max(\bar{g}) = 0.05$$

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

When $b_{\text{min}} = 0$, $T = 400$, $N = 100$ (48 sec)

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

but it varies per run!

$b_{\text{min}} = 50$, $T = 400$, $N = 100$

$$0.00018217 \quad 1.8037e-07$$

and still varies

$b_{\text{min}} = 1600$, $T = 400$, $N = 100$ (48 sec)

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

vary still.

$b_{\text{min}} = 0$, $T = 2000$, $N = 100$ (150 sec)

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

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

$$2.6731e-05 \quad N=1 \quad (2 \text{ sec})$$

$$9.6657e-05$$

$$1.511e-05$$

$$1.0546e-05$$

3 obs: 1) b_{min} doesn't matter (why? b/c doesn't impact FER)

2) longer T decreases \bar{g}^* } Ryan's conjecture was
3) decreases dispersion across n } right.

Moving on to MSE:

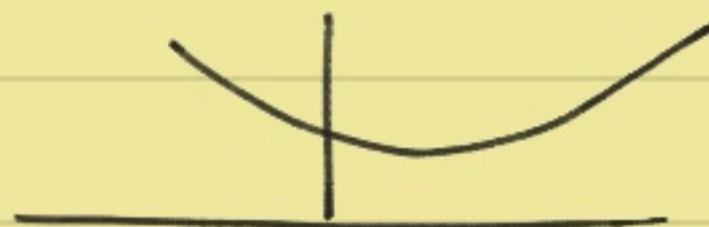
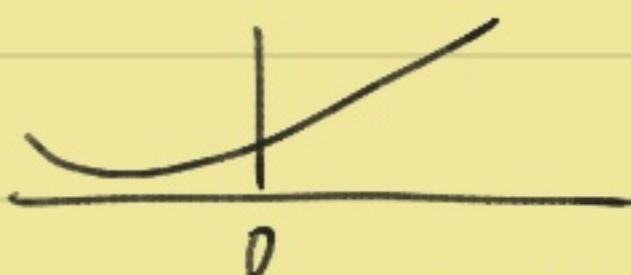
Note $\text{MSE}(y_j) = \frac{1}{T} \sum_{t=1}^T (y_{j,t} - \hat{y}_{j,t})^2$ (Brand & Gans 2006)

$\text{MSE} = \text{FEV}$ b/c $\text{FE} = x_t - x_{t-1}$

$\text{Var(FE)} = E[\text{FE}^2] = \text{mean squared error}$

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

\Rightarrow 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 $\Sigma = \text{VC}(\text{shocks})$) the lower \bar{g}^* (more negative)

Ok so what Roger 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 forecasts, but not Y . What's the \bar{g} that mins the FEV?

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

whereas for my exercise I get $\bar{g} = 2.424e-05$ $1.7218e-8$
($b_{\min} = 0$, $T = 2000$, $N = 100$)

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 FER 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 action 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!

Schenkman & Xiong, 2003 "Overconfidence & Speculative bubbles"
(setup on optimism & pessimism on financial markets)
→ This stuff relates to my confirmation bias idea

Understand why minFEV

23 Nov 2015

Ryan's method: $y = \text{generated by gain} = 0.145$

Now set $K = \bar{g}^* = \arg\min \text{FEV}(y(\bar{g} = 0.145))$

My method: $K = \bar{g} = \arg\min \text{FEV}(y(\bar{g}))$

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 $\bar{g} < \bar{g}^{\text{DGP}}$. I don't think this is

a general statement; i.e. I think he only expected that compared to $\bar{g}^{\text{COMP}} = 0.145$. Why do I think this?

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

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

- So the statement is that even as data g is generated w/ $k_t = \bar{g}^{\text{COMP}}$, that leads to FEs b/c the PLM doesn't coincide w/ the ALM. So if $k_t = \bar{g}^{\text{COMP}}$ and FEs switch sign, that means that by lowering \bar{g} , we could lower FE b/c agents are overpredicting.
- 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 $k = \bar{g}^{\text{COMP}}$. If $k_t = 0$, and initialized at RE, $\hat{E}_t \pi_{t+1} = 0 + b_1 s_t = \text{RE first}$. 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, $b^* = 0$

Cannot be optimal.

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

What does it mean if this is < 0 ?

$$\text{fcsks are } \hat{E}[\bar{\pi}_{t+1}] = \bar{\pi}_{t-1} + b_1 s_t$$

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

If $k < 0$, then if I underestimated π_t ($\text{test} < \pi_t$)

then I lower $\bar{\pi}$, that is, I lower my fcsk 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 gives 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 these:



But: neither my nor Ryan's should give $k^* < 0$.

So there's still 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) $\bar{g}^* \rightarrow 0$ from below 2) shrinks $\text{Var}(\bar{g}_n^*)$

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 centered around \pm some number, k^* 's become positive!

2) Changing Σ doesn't really do anything.

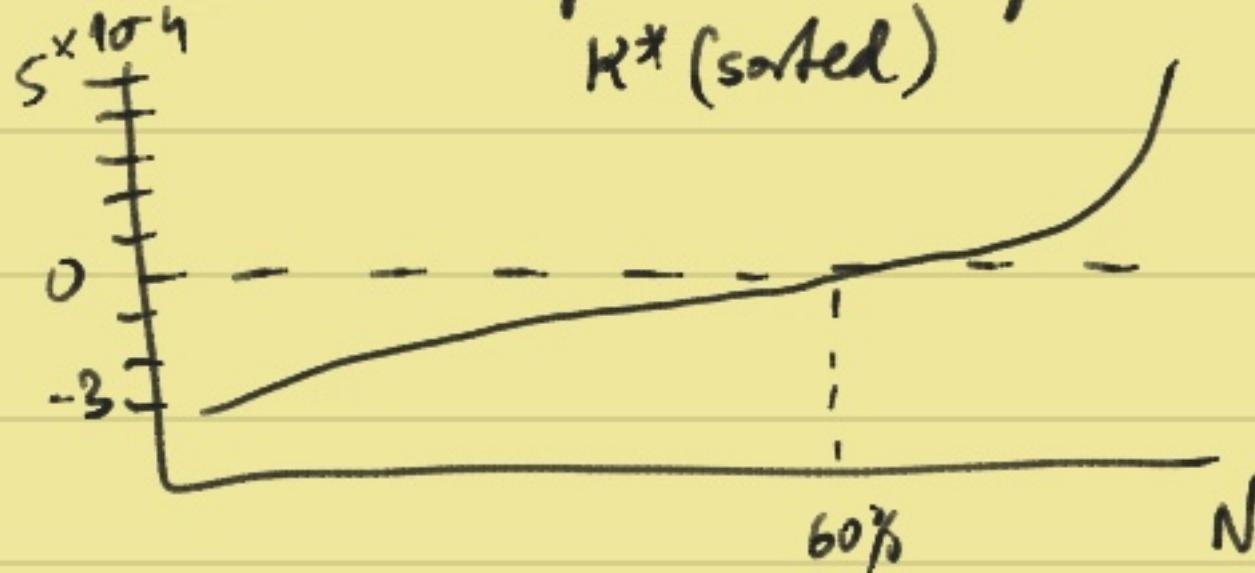
Why is this happening? And why doesn't the sign(mean(shocks)) matter? (\rightarrow i.e. why doesn't it matter whether shocks are centered around 1 or -1?)

\rightarrow Try shutting off all shocks except 1 \rightarrow 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 $T \uparrow$, $k^* \rightarrow 0$ but always from below

and the distnb of k^* always is skewed like this:



Reading Berkoo 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 (abstact shocks) given these beliefs?

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

The E-stability condition is the differential equation

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

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

In both cases w/o observables)

Camp's criterion $\theta_t = \text{PLM} - E(\text{ALM})$, so it's really

the E-stability diff. equation $\theta_t = \frac{d}{dt}(\phi) \quad (!)$

Susanto meeting

2 Dec 2019

Work after

3 Dec 2019

Ball 1994 AER

Deflation causes recessions as we thought.

But disinflation (changing the growth of money \downarrow)
causes a boom.

I think the idea is this:

$$x_t = m_t - E_t m_{t+1} \quad \text{(roughly)}$$

↑ all future m^s

↓ deflation causes $x_t \downarrow$

but disinflation is $m \downarrow$ 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!

$$x_t = -\beta \pi_t \pi_t + \sum_{T=1}^{\infty} \beta^{T-t} \left((1-\beta)x_{T+1} + \underbrace{2(1-\beta k_T)}_{<0} \pi_{T+1} \right)$$

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 $E\pi_{t+1} \downarrow$,

$\hat{q}_t \uparrow$ 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 bc they anticipate 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 UK model.

One thing I'm a little worried about is that $E(\cdot)$ moving strengthens the Ball effect, and $E(\cdot)$ 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 $E(\cdot)$ to move. I can get them to move otherwise.

Instrument instability seems to happen 4 Dec 2019

here in the sense that expectations are the instrument.

In particular, the FE oscillates.

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

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

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

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

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

$$FE_{t-1} = \pi_t - b_1 s_{t-1} - \bar{\pi}_0 - k^{-1} \sum_{s=0}^{t-2} FE_s \quad | \pm FE_j \\ j=t-2, \dots, 1''$$

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

Ignore this 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

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

If we only had one lag, this would be

$$kx + 1 = 0$$

$$x = -k^{-1} > -1$$

That would be stable
but oscillating

Two lags

$$kx^2 + x + 1 = 0$$

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

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

For the roots to be real $1 - 4k > 0 \Rightarrow 1 > 4k$

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

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

\rightarrow 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, but 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 γ_{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 γ_{π} must come via the role of i in the rule x .

$$\left. \begin{array}{l} x = -\beta i_t + E(\text{stuff}) \\ \pi_t = \kappa x_t + E(\text{stuff}) \\ i_t = \gamma_{\pi} \pi_t \end{array} \right\} \begin{array}{l} \text{In period 1, a shock hits, } E \\ \text{moves a little, } (x, \pi) \text{ move, } i \\ \text{moves.} \end{array}$$

Evening of period 1: $E(\cdot)$ adjusts. 2 things influence the adjustment: 1) $\kappa = \bar{g}$ the size 2) γ_{π} the size of Ball's disinflationary boom effect (direction of E -adjustment)

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

↑ ↑
↑ governs the change here

γ_π governs the change in both, and the bigger it is, the more $E(.)$ moves compared to π_t , opening up FEs.

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

Chaining ; works partly via its effect on x_t and π_t today, but its main effect is on $E(.)$.

So in that sense it's like instrument instability b/c the instrument becomes unstable, but: it's also not like instr. instab. b/c a too high γ_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 \rightarrow so it's not quite the case that empirical IRFs never oscillate ...
- L'Hourdis estimated $\hat{\gamma}_\pi \approx 1.004$ or sthg
Note: it's Blanchard, L'Huillier, Lourenco, Mers & Nauk, AER 2013
and they get $\hat{\gamma}_\pi = 1.0137$, $\hat{\gamma}_x = 0.005$
- Try $\hat{\gamma}_\pi < 1$ \rightarrow it seems like if for a fwd-looking system $\hat{\gamma}_\pi > 1$ gives stability, for a bw-looking one, $\hat{\gamma}_\pi < 1$ will.
- Consider flex price model

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

- 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.

Work after

5 Dec 2019

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 p_i , the persistence of the non pol shock, 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 FE's

→ similarly corr errors in the same direction

Peter Ireland similar paper to Ball:

Stoppage Inflation, B.g & Small 1997 JMCB

Ric. EA 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

$E(\cdot)$ and then for observables.