MOVING THE NEXT NOTES to the iPad. 10 August 2020

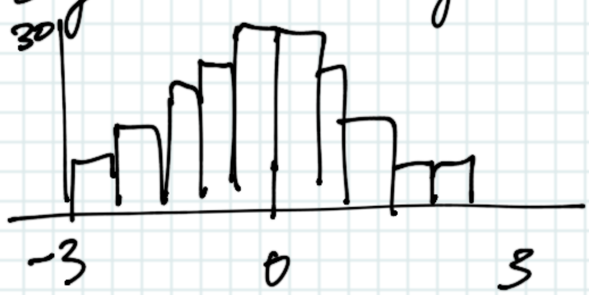
So: trying to generate truth w/ more action in expectations (R, b).

Observed:

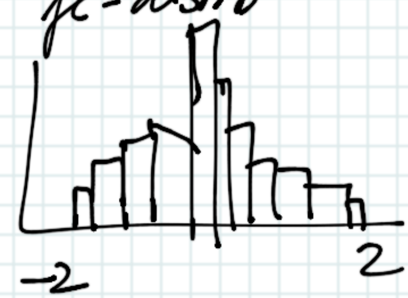
1) If you increase the fe-support, fe's in the simulation become more tight around zero, and extrema are smaller. (E.g. $fe \in (-2,2)$ vs. $fe \in (-0.5, 0.5)$).

2) If you scale up α's, fe-distribution becomes more spread out: fatter tails

E.g. $4 \cdot \alpha^{true}$ gives the following fe-distrib



instead of sthg like



and $\bar{\bar{a}}$ fluctuates much more.

Estimating 4.α

- loss was initially much smaller 4.95 instead of on the order of 4000.
- estimation gets a lot of "fe was nan" messages and also a new one: "obj function returned NaN, trying a new point". then...
  - ↳ I already got this message 9 times
- It took 585 sec.
- implied k^{-1} is <0 4 times!
- → maybe that's for negative k ?
- Again, $loss(\hat{\alpha}) = 3.2 < loss(\alpha^{true}) = 416$.
- Not possible to evaluate loss function?!
  - ↳ It seems to be nans everywhere?
    Is that why the loss fct returned nan b/c all simulations were nan in the cross-section?
- Look at the loss: it sucks:
  - for α_1, α_5, it's apparently always nan
  - for α_2, α_4 it blows up
  - for α_3, it's oscillatory.

Any $\alpha > 0.5$ gives all nans. Maybe that's b/c since the truth is higher, it more quickly explodes?

- I'm not thrilled that many histories explode.
- Also very many VAR instability problems.

Ld's do more or less peak at the truth even now. (well, α_3 doesn't and α_1 & α_5 have 2 minima around the true value, but still)

→ Does this indicate that the data-generation is wrong or the estimation?

- True moments changed very much:
  - moments pertaining to x scaled up, ones pertaining to π or i scaled down / were unchanged
  - own moments or cross-moments w/ π or i have the same shape
  - cross-moments w/ x have changed shape: are now more U- or reverse U-shaped.

↳ I think these make sense b/c since k is low, and $i = 1.5\pi$, x is the one that absorbs movements in expectations, while the pass-thru to π and to i is small.

⇒ The data-generation I think is correct. But this kinda suggests that the mean moments are also (more or loss) computed correctly, of which I'm not convinced.

↳ Let's try an estim w/ $\psi_x = 0.5$.

- But suppose the data is generated correctly. Why isn't the "x absorbs all" mechanism matchable by the estimation? I mean, moments didn't move far from initial ones.

    ↳ Explosion % reveals that in the estimation, either 0% or 100% of the N simulations explode. No, sometimes it's 19, 79 or 99%. And it seems that the obj fct returns 'nan' when 100% explode. Yes.

    I learn also, by the way, that each iter of the solver evaluates the loss function exactly 6 times.

Ask Ryan

↳ Why? It's not changing α, or if it is, only one and only marginally. I'm also noticing that it's only considering α's really close to α_0 ⇒ multistart!

I'm also noticing that the simulation is not giving the 'fe was nan' error most of the time, even though the solver gets a positive expl-percent. Hm!

When the "fe was nan" error comes, then the explosion-counter does catch it. But apparently there are a bunch of explosions that come not from these.

Ok, it's clear: when fe is nan, it's always caught. The other thing that causes "explosive" simulations is $k^{-1} < 0$. These get caught and register as explosions, but aren't really.

→ My check for "global nonnegativity" one wasn't catching everything!

Iter 40 : that's where the troubles start.

⇒ Initially, fe is nan (maybe the gain is too high) but then the bulk of the supposed explosions come from $k < 0$.

Ok... even when fe is nan it seems it's b/c or at the same time as $k < 0$ ...

→ It's at the same time. The initial iteration involves 5% "fe was nan" (5 / 100 simulations) but 100% have $k < 0$ at one point!

Let's see for the July 6 dataset ... and yes,
the initial iters involve 1% fe nan
69% k < 0
and later iters have 2% fe nan
86% k < 0

Converges after 23 iter.
Thus no more problems after iter 4.

Why is fe often nan early in the search and not late? Is it b/c of the initial values α_0?
And could a finer, and more broad fe-grid for the global nonnegativity test work?
↳ I introduced the variable broaden=2,
! which extends the fegrid-fine by ± broaden at both ends. It seems like this catches & subsumes all the previous errors!

It still finds the same min though :S
(But at least it's 2/3 of the time)

If I make the loss fct = nan when k<0 globally, I can engineer the "Obj fct returned nan" error message.

Does this influence the estimate? (Previously I had set the objective to 1e+10) No it doesn't.
→ I don't think so but it does affect the plots of the loss I would think!

Let's go back to the scaled-up truth and plot the loss
should understand why $k<0$: is it, as I think, that α isn't convex and so a fc outside the grid is extrapolated to a $k<0$?

The other thing is that this doesn't seem to really affect the behavior of the loss fct. It still converges to the same (wrong) thing. Actually, for the 4.α truth, it converges to sthg different, but most likely that's b/c it took an avg of a smaller set.

The loss: $k<0$ and fc = nan return. Does that mean that fcgrid-fine needs to be even broader?
Not necessarily b/c now it happens that e.g.

fc = nan in 14%

but $k<0$ in 0%.

In fact, $k<0$ doesn't happen a lot! fc is the problem here!

Why is it that for the estimation, $k \approx 0$ is the root of all evil, but in plotting the loss, suddenly $fe = \text{nan}$ becomes the problem #1? Is it that the estim avoids high α's (or maybe also very low ones) that could cause turmoil?

Materials 40, Section 4: Look into behavior of simulation. 11 August 2020

I think I learn a lot from this. Fig 7, which shows means & distribs of k^{-1}, $\bar{\pi}$ and fe are indicative.

1) While k^{-1} spends a lot of time being near 0 (see histogram), mean$(k^{-1}) \approx 0.0185$ (see means)
   ↳ If the moments capture the latter aspect, then the estim should have a hard time pushing α_3 down. But the moments should more capture the first. Moreover, what does it matter if mean$(k^{-1}) = 0.019$? It still can enter the lower parts of its state space, and it does! So scrap that.

2) The cross-sectional mean of $\bar{\pi}$ & of fe ↓ as N↑. Maybe this is why loss↑ in N? I'm not sure this should happen? ⟶

As $N \to \infty$, $e \to 0$. So $fe \to 0$ and thus $\bar{\bar{a}}$ too.

But that means that mean moments will reflect a sim in which $\bar{\bar{a}}$ isn't moving much, and fe are very small.

Now turn to changing α's: do the following exercises:

- $4 \cdot \alpha$
- More edges: α_1 & α_5
- More middle: α_2 & α_4
- More 0-point: α_3

Obs. 1: If all α's ≤ 0.1, simulation doesn't blow up. At 0.11, they do. (2 occurrences of "fe was nan", one of which was also $k_{-t-1} < 0$)

Obs. 2: $\alpha = 0$ doesn't cause explosions.

Obs 3: In the early scenarios, "fe was nan" is the only message. In the late ones, it comes together w/ "$k_{t-1} < 0$". Does this mean that $k < 0$ when α_3 is large?

↳ Making the code output α suggests that:

- "fe was nan" alone occurs when α_1 & $\alpha_5 \geq 0.11$
- "fe was nan" + "$k_{t-1} < 0$" occurs when α_2 & $\alpha_4 \geq 0.11$

- α_3 doesn't shit any water: it can also be 0.5, that doesn't cause explosions. (0-neighborhood indifference problem)
- these seem to occur across shock histories.

Now let's look at the plots of the means:

Scenario 1: Varying α_1 and α_5: 0, 0.05, 0.1

- Shifts k^{-1} up from (0.01, 0.012) to (0.02, 0.03)
- Shifts $\bar{\pi}$ from (−0.01, 0.01) to (−0.02, 0.04)
- Barely shifts fe: it remains in the (−0.2, 0.2) range

Scenario 2: Varying α_2 and α_4

- Shifts k^{-1} up from (0.005, 0.001) to (0.05, 0.06)
- Barely shifts $\bar{\pi}$: it remains in the (−0.01, 0.03) range
- Barely shifts fe: it remains in the (−0.2, 0.2) range

Scenario 3: Varying α_3:

- Shifts k^{-1} up from (0.015, 0.02) to (0.05, 0.06)
- Barely shifts $\bar{\pi}$: it remains in the (−0.01, 0.03) range
- Barely shifts fe: it remains in the (−0.2, 0.2) range

↳ So: why aren't $\alpha_{2,3,4}$ moving $\bar{\pi}$ & fe?
  ↳ is it b/c in that range $k^{-1} \cdot fe$ is a really small number? I.e.: is the 0-neighborhood indifference a problem

here too?

→ I think so! I think what is happening is that the entire $k \in (-1, 1)$-region yields too small $k^{-1} \cdot k$ products such that they do matter a little bit (the loss does take a min there) but not a lot (the loss is very flat)

⇒ To be identified, you need to consider α's associated w/ $k > 1$ (possibly even greater). But: you also need a truth that's not too big, b/c $\alpha > 0.1$ seems to be ill tolerated by the model.

→ TO DO!

For Peter meeting:

① Loss does have a min, almost at right spot. (Fig 1)

a) Need convexity assumption – don't get removal (Fig. 2)

b) Rescaling W affects loss, although no indication of inversion issues. (Fig 3.b)

c) Adding E(·) screws things up in a way I don't get. (Fig 4)

(② Truth w/ more action in E(·)) (Fig 6)

I thought that if E(·) aren't moving, then both rescaling

and informativeness of moments might be screwed up.
Explosions hinted at something else.
$\alpha \overset{!}{\leq} 0.1$ to avoid explosions.

(3) Behavior of simulation

a) Loss ↑ in N ↑ b/c $\bar{\pi}$ & fe ↓ in N? (Fig 7 b&d)

b) Fig 8: Only α_1 & α_5 can affect $\bar{\pi}$, and even that can't affect fe. (Fig 8)

Peter meeting

11 Aug 2020

Fig 1. Loss higher at $\hat{\alpha}$ vs α^{true}

Again: b/c data from model is nonlinear, this may not suggest that α^{true} is a local min.

Thinks that a lot of problems come from that est a nonstat model is tricky.

The next issue: holding α's fixed, but estimating more than one α may not work. This would happen if you est $Y = \beta_1 X_1 + \beta_2 X_2$ and X_1 & X_2 multicollinear.

↳ might be some weird interaction between α's so that the concavity restriction helps provide curvature.

Fig 8 conclusions sound exactly right.

Rescaling: still quite strange

Fix α, check loss w/ and w/o scaling → should recover the scaling factor.

Also explains why imposing convexity helps.

Wanna guard against explaining what is really a coding error. → Should go back to that.

The level of loss fct has no meaning; only the curvature.

Min in Fig 1. don't line up w/ truth: not a problem b/c if you sim using a diff c, you'd get a diff min.

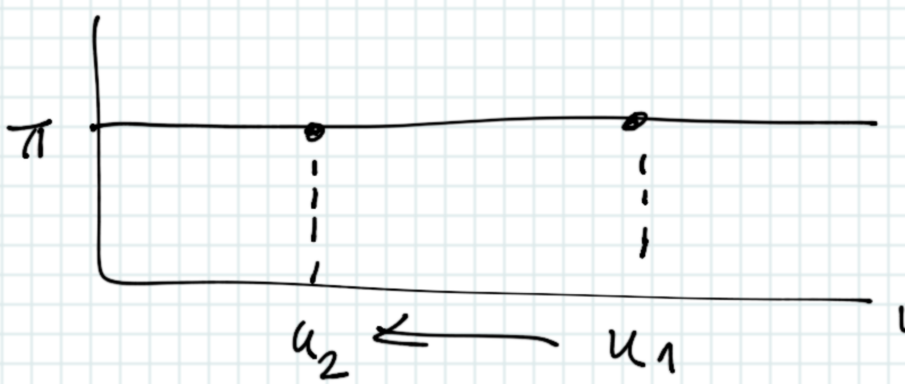
If std. errs large, can be b/c of
- → ID → doesn't go away as $N \to \infty$
- → sampling error
  - ↳ goes away as $N \to \infty$

Recall the logistic functional specification for anchoring fct: somehow more structure has to be imposed on the anchoring function to guide the info in the data

Analogy: Selena?
Silvana Tenreyro: $\hat{PC}$: slope is ≈ 0 $\rightarrow$ b/c CB is credible! So est a PC w/ these data doesn't tell you what would happen if CB abandoned target. (b/c that's just not in the data!)

$\rightarrow$ i.e. obs that have modest fe, we don't learn a lot about how agents learn from forecast errors.

"flat PC": "π isn't moving in tandem w/ u anymore"



$\hookrightarrow$ he says that the "flat PC" term refers to the idea that π doesn't follow u, not the other way around!

low u doesn't lead to any diff π!

- A fall-back approach: a simpler functional form for anch. fct.
- Or: pick unconditional moments w/o VAR or weighting matrix

He called it an in-between btw est & calibration.
"Moment-matching": take sample moments & match those.

Successful esti will involve picking restrictions on 1st OR on the anchoring fct to allow to glean info that is in the data.

↳ And I'm not sure why he calls the "moment-matching" by that name: it's not much different than my SMM.

Let's try the thing w/ α's out in the edges.

12 August 2020

Attempt ①: July 6 data, fegrid = [-2, -1.5, 0, 1.5, 2]
(truth same) converged, not much change.

Attempt ②: July 6 data, fegrid = [-4, -3, 0, 3, 4]

Took 500 sec & converged to sthg quite diff! Not correct but maybe getting near?

Attempt ③: July 6 data, fegrid = [-4, -3 ; 3, 4] nfe=4
Took 115 sec & converged to sthg very diff. Isn't correct either but the thing is that the "truth" doesn't have a lot of fe in the 3-4% range.

Attempt ④: July 6 data, fegrid = [-4, -3, 0, 3, 4]
but w/ Wmid=1000.
Took 233 sec. Looks good. fe ±3 & 4 are still off, but I'm hopeful.

Attempt (5): Generate truth with large forecast errors which you can est. w/ fegrid = [-4, -3, 0, 3, 4]

First day Aug 10 data ($4 \cdot \alpha^{true}$)

(w/ Wmid = 2000)

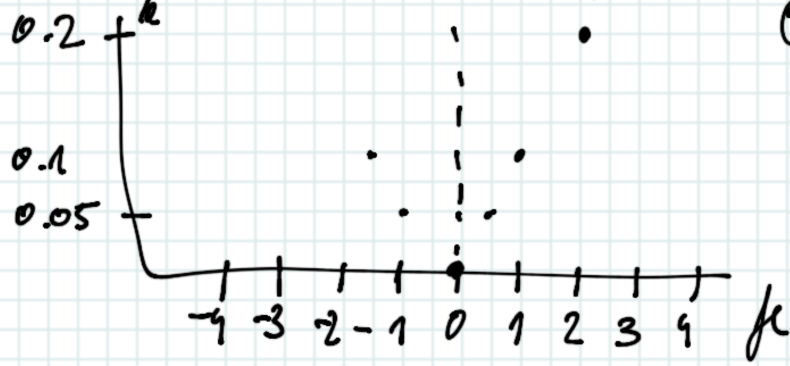
I'm also noting that the convexity thingy ain't working well.

Took 643 sec.

Got a bunch of "3% of sim histories exploded" messages.

There we go, lookin' good!

~~Attempt~~ (6): Make α's associated w/ small fe large, so the truth looks something like:



Oh shit. This is exactly the 4 x times thing!

And $5 \times \alpha^{true}$ explodes already.

I don't think I can generate a dataset w/ bigger fe b/c large fe need large α's, but sufficiently large α's also explode, so...

Real data with Attempt (5) settings leads to a 1-2% sim histories exploded sim. 390 sec.

I think I'm going to leave it at that and consider the rescaling issue. Can I recover the scaling factor? Aand... no. Actually yes, if you recall that the loss involves squaring W.

↳ Ask Ryan: could it make a diff that in Lsqnonlin you specify the equations resids · W?

<u>Applied Young Economists Webinar</u> (AYEW) <u>22 Aug 2022</u>

Nan Hu & Zexi Sun "Vague Talking at Central Bank's Press Conference: News or Noise?" (19 participants)
Hong Kong Institute for Monetary and Financial Research (← note)

interpret as "news" vs. noise
Vague talking that includes words like "risk", "uncertainty", "volatility" and "turbulence" raises stock returns
→ b/c investors expect expansionary mon pol in the future: the ⊕ news dominate the ⊖ news that "current situ is bad".
Expansionary disinflation (Ball 1994)

dictionary from Loughran & ..., freq. of uncertain words

Regression: $R_t = b_0 + b_1 \Delta \text{vaguetalking}_t + \gamma_b' X_t + \varepsilon_T$
First diff b/c EMH: stocks should only react to <u>surprise</u>

Prepare Ryan meeting:

(1) Loss has a min at truth (Fig 1.)

(2) Loss is indifferent to α's such that $|fe| \leq 1$ (Fig. 8)

(3) Having a more action in fe truth and est-ing α's corresponding to large fe values (Fig 15)

Details & problems:

(4) Solver evaluates loss 6 times for the same α?

(5) Rescaling: changes shape of loss while no sign of inversion issue?

<u>Ryan meeting</u> <u>12 August 2020</u>

- Need to est σ of shocks in order to get model-implied moments closer to those of the data, at least for π. → or play around w/ 'em.
  ↳ If shocks too small, no fe ⇒ ID problems

- Isn't the whole story, b/c edgepoints also off

- Fig 8: would learn more if you plotted the auto-covariograms for these exercises!

→ So what's still amiss? We still don't know.

Weighting matrix: Eq (2): multiply by $\sqrt{diag(w)}$

→ Take the $\sqrt{\cdot}$ b/c otherwise you may put all weight on one moment.

· My rescaling strategy likely not to solve inversion issues b/c it's the ratio $\frac{\max(\text{element})}{\min(\text{element})}$ of elements of Σ which matters. If it's more than 10e6, then problem.

↳ Ryan's strategy: Fix largest el. of Σ and fix all elements to be not smaller than 1/10 000 the largest.

• Not aligning the right els of W w/ the right moments could be a potential problem.

• Search alg:

6 evals b/c it computes gradient & jacobian.

(by 100 say) You can multiply the step for eval of ↑, then the est. of derivatives is less precise but you get a better idea of the slope.

"first/fwd difference" or othg in matlab

Search algorithm evaluates loss 6 times for each guess b/c it 1.) evaluates at the guess

2.) changes each x_i a LITTLE bit to estimate first and second derivatives. It changes one α_i at a time, so this gives 5 additional evaluations.

Main point: the story that we can't use small fe to learn about how agents learn from fe is good, but it isn't the whole story. And one may be that shocks are too small, partly responsible for why there are no large fe in the data.

<u>Work after</u> <u>13 Aug 2020</u>

I have 2 open businesses:

1) The ID question: - larger fe needed in sample
$\rightarrow \mathcal{Z}^{\text{shocks}}$

2) The weighting matrix: - take sqrt(·) of els
- figure out rescaling

Point 2): The weighting matrix.

Taking sqrt of the elements of W does change the est results. But in a sense these results are kinda better than the flat default Nsimul.

For the "α's in the edges" specification $\sqrt{W}$ doesn't matter a lot. I guess not b/c that one is better ID-ed anyhow!

What I notice though is that taking $\sqrt{W}$, I <u>don't</u> get $L(\sqrt{W}) = L(W^{rescale}) \cdot$ scaling factor (that I don't get $L(W) = L(W^{rescale}) \cdot scf^2$ is no surprise)

Ok but what does hold is:

$$L(\sqrt{W}) = L(\sqrt{W^{rescale}}) \cdot scf.$$ So it still works, as it should.

The thing is: maybe, w/ larger shocks, $Var(E(\cdot))$ will be higher so the weighting matrix won't be an issue at all. Still: I need to understand why the weighting matrix rescaling changes the loss fct.

Comparing to the default Nsimul picture, does $\sqrt{W}$ change the shape of the loss? → It does, a little, but not a lot. This is what I expected, given that it weights diff. moments and therefore also obtains diff. estimates.

Now scale W and see if the loss changes.

It does — but again only for α_1 & α_5 ...

I've also checked what Ryan said about the order being screwed up: that doesn't seem to be happening either.

there are 2 options: either I'm somehow rescaling in a way that does change relationships, or there is something else somewhere.

I think it's Option 2 b/c $\frac{max(W)}{min(W)} = \frac{max(W^{rescale})}{min(W^{rescale})}$

also we should have

$$W^{rescale} = \frac{1}{scl} W$$ and that holds too!

↳ I'm really starting to think that the change in the shape of the loss fct has sthg weird to do w/ the indifference of $\alpha_2, \alpha_3, \alpha_4$.

The rescaling makes the loss smaller → maybe the non-rescaled one would have the same shape if I zoomed in more?

Erm... not quite. Can it be that it's a numerical thingy that it changes shape?

Honestly, I don't think so.

Try to set $W = I$. Now rescale loss and see if it changes.

→ it does! And the same way too!

I have an idea. maybe the loss fct is changing shape b/c the convexity moment is in place.

Shut off Wdiffs2 → Wdiffs2 = 0.

Set $W = I$ and compare loss w/ & w/o rescaling!

⇒ They have the same shape!

Redo w/ $W = \Sigma^{-1}$ & Wdiffs2 = 0 ⇒ same thing!

YEAH! That was it! The convexity moment was effectively relatively weighted more when I scaled down the weights on the others!

I also don't get why evaluating the loss fct now takes longer than before. Producing these loss plots takes 95s. instead of 75 sec.

→ today

→ Whoa, I can't believe that it was that simple! So I can move on to what matters: the forecast errors and the variance of shocks. Let's think: how best to do this:

A first pass thing is to see whether simulated fe ↑ as σ↑. (command_check_simulation_approx.m)

Set $N=1$ (rng(1)). Default $b_z = 1$ $z = r, u, i$

Default: $fe \in (-2,2)$. $(-2,2)$ $\alpha = 2.5 \cdot \alpha^{true}$.

$(-5,5)$ $\alpha = 4 \cdot \alpha^{true}$ $\alpha = \alpha^{true}$

Raise $b_u = 2 \rightarrow$ explodes.

Raise $b_u = 1.2$ $\Rightarrow fe \in (-4,4)$. explodes $(-2,2)$

Raise $b_u = 1.5$ $\Rightarrow fe \in (-5,5)$. —"— $(-4,4)$

$b_u = 4 \rightarrow (-10,10)$ $b_u = 1.8 \rightarrow$ explodes. —"— $(-5,5)$

Raise $b_i = 1.2$ $\Rightarrow fe \in (-2,2)$ $(-2,2)$ $(-5,5)$ ($b_i = 2$ too)

$b_i = 4$ $\Rightarrow$ same. $(-2,2)$ explodes

b_i should have an effect though.

$b_i = 8$ $\Rightarrow fe \in (-4,4)$ $(-2,2)$ $\rightarrow$ It does, just a much smaller one.

Raise $b_r = 8$ $\Rightarrow fe \in (-5,5)$ $(-4,4)$ $b_r = 2$ explodes,

↳ Also a much smaller effect. I'm a little surprised that b_r apparently has a higher effect than b_i b/c ... oh no, idiot... of course b_r has a higher effect b/c the int. rate works thru the IS-curve!

$$\pi = \kappa x + u$$

$$x = -b_i + r^n$$

$$i = \psi_\pi \pi_t + \bar{i}$$

So the strength of the effect is of course $b_u > b_r > b_i$.

Ok, so that was like initial reconnoitering. I guess the next step is going to be to see how the estimation reacts to 1) true data being generated via higher τ. Maybe the easiest is to do $\tau_u = 4$ as a first pass. w/ $\alpha^{true} = \alpha^{true}$ (0.05, 0.025, 0, -11-, -11-)

this is the Aug 13 data!

Shit. VAR complaints, a lot of them.

Ridge? No complaints.

But then maybe the estimation needs to use ridge too.

Yep, it definitely complains.

Ok now it has some explosions (fe was nan) and it still has VAR issues despite the ridge. Let's try to lower (?) the ridge parameter λ.

(It was 0.001, now it's 0.0001)

immediate VAR error.

So raise it to 0.01

seemed to work well, but then an "input to rank must not contain NaN or Inf" error. Hmm.

Let's try $\tau_u = 2$. Think I need to rethink the ridge?

Yeah, immediate error. Can create data w/o ridge? No.

Ok, so the situ is the following: for higher b, data creation and estimation needs ridge. So TO DO: look into ridge and confirm that it's correct. Create some replicatable simple example.

Confirming ridge 17 Aug. 2020

The issue of centering & scaling: Matlab's default for ridge is to center & scale: X is standardized (mean 0, var 1) and y is centered (demeaned I guess?)
There is mention of this on some websites where I check ridge. Why is this important?

Statweb.stanford.edu
Ridge is related to PCA: sort of weighted X's are j^{th} principal components of X.
→ Ridge projects y on the PC's w/ large variance.

Statlect.com provides a hint as to why:
OLS is scale-invariant: β^{OLS} of $R \cdot X = R^{-1} \cdot \beta^{OLS}$ of X.
β^{ridge} is NOT scale-invariant! Therefore, whether you express variables in m or cm, e.g., matters!
↳ We therefore standardize all variables in ridge to

avoid undesirable scaling effects.

But I think you only need to stdize X, not Y?

Matlab helpfile is confusing me now:

$$B0 = \left[\frac{\text{mean}(Y) - \text{mean}(X) \cdot B1}{\text{std}(X,0,1)}, \frac{B1}{\text{std}(X,0,1)} \right]$$

Maybe Matlab is actually standardizing in each case, it just computes $B0$ as above?

Stata notes on Ridge (ncss-wpengine.netdna-ssl.com)

- standardize all variables: X and Y
- when final results are presented, they are adjusted back into their original scale. (Matlab's "0").

→ When "scaled = 0" "ridge restores the coefficients to the scale of the original data".

↳ It's out: · Matlab standardizes in any case
· scale = 1 or 0 determines whether Matlab scales them back to their original scale.

→ So let's stdize Y & X!

Many observations:

1) Still not the same as Matlab.

2) Matlab's ridge.m doesn't seem to be scaling Y.

↳ If I use the original, non-stdized Y, I get the same as Matlab's Ridge, scale = 1.

✓ ok, but should one std-ize Y or not?

Let's see:

$$\hat{\beta}^{OLS} = (X'X)^{-1} X'Y. \qquad \text{Let } Z = XR \quad \text{rescaler.}$$

$$\tilde{\beta}^{OLS} = ((XR)' XR)^{-1} (ZR)'Y$$

$$= (R'X'XR)^{-1} R'X'Y$$

$$= (R')^{-1} (X'X)^{-1} (R^{-1} R') X'Y$$

$$= (R')^{-1} (X'X)^{-1} X'Y \quad = R'^{-1} \hat{\beta}^{OLS}.$$

$$\hat{\beta}^{ridge} = (X'X + \lambda I)^{-1} X'Y \qquad \text{Again } Z = XR$$

$$\tilde{\beta}^{ridge} = ((XR)' XR + \lambda I)^{-1} (XR)'Y$$

$$= (R'X'XR + \lambda I)^{-1} R'X'Y$$

$$= (R'X'XR + \lambda R'(R')^{-1} R^{-1} R)^{-1} R'X'Y \quad \Big| \text{ Factor } R' \text{ from left and } R \text{ from right}$$

$$= [R' (X'X + \lambda (R')^{-1} R^{-1}) R]^{-1} R'XY$$

$$= R^{-1} (X'X + \lambda (R')^{-1} R^{-1})^{-1} (R')^{-1} R'XY$$

↑ inverse of product switches order.

$$\tilde{\beta}^{ridge} = R^{-1}\left(X'X + \lambda(R')^{-1}R^{-1}\right)^{-1}(R')^{-1}R'X'Y$$

$$= R^{-1}\left(X'X + \lambda \underline{(R')^{-1}R^{-1}}\right)^{-1} X'Y$$

For this to be equal to $\hat{\beta}^{ridge}$, we need $(R')^{-1}R^{-1} = I$ i.e. R needs to be orthonormal. Generally doesn't hold!

↳ only need to standardize X! (Not Y.)

Do need to standardize but also need to scale back! → Nice stats.stackexchange question & answer: (Philipp Burkhardt) For when only regressors std-ized.

18 Aug 2020

Let $z_j := \dfrac{x_j - \bar{x}_j}{s_j}$ standardized regressor j

Then we have the regression line

$$E[Y] = \beta_0 + \sum_{j=1}^{k} \beta_j z_j$$

which gives fitted values

$$\hat{Y} = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j z_j \, . \quad \text{Write this as}$$

$$\hat{Y} = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j \left(\frac{x_j - \bar{x}_j}{s_j}\right) \quad | \text{ Rearrange}$$

Matlab has $\bar{Y}$ in here too.

$$\hat{Y} = \underbrace{\left(\hat{\beta}_0 - \sum_{j=1}^{k} \hat{\beta}_j \frac{\bar{x}_j}{s_j}\right)}_{\text{Scaled-back intercept}} + \sum_{j=1}^{k} \frac{\hat{\beta}_j}{s_j} x_j$$

β scaled back (= what Matlab has)

If both regressor and regressand were scaled:

$$\hat{Y}_{scaled} = \frac{\hat{Y}_{unscaled} - \bar{y}}{S_y} = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j \left(\frac{x_j - \bar{x}_j}{S_j} \right) \Big| \cdot S_y + \bar{y}$$

$$\hat{Y}_{unscaled} = \bar{y} + S_y \hat{\beta}_0 + S_y \sum_{j=1}^{k} \hat{\beta}_j \left(\frac{x_j - \bar{x}_j}{S_j} \right)$$

$$= \Big(\bar{y} + S_y \underbrace{\hat{\beta}_0}_{=0} - S_y \sum_{j=1}^{k} \frac{\hat{\beta}_j}{S_j} \bar{x}_j \Big) + S_y \sum_{j=1}^{k} \frac{\hat{\beta}_j}{S_j} x_j$$

This doesn't correspond to Matlab either : extra terms absent in Matlab.

I don't think Matlab is scaling Y. I suspect mean(Y) is added back to take care of E(Y) somehow?

Actually, Matlab's help for ridge.m, "Coefficient Scaling" explains it. Matlab standardizes X and centers Y (uses $\hat{Y} := Y - \text{mean}(Y)$), which is why mean(y) shows up in the intercept.

Ok: at least whether you use demeaned Y or not doesn't matter for $\beta_1, \dots, \beta_k$, only for β_0, the intercept.

Ha! If you don't use demeaned Y, & scale back w/o $\bar{y}$, you get the same thing as if you used demeaned $\bar{y}$ and scaled back w/ $\bar{y}$.

Ah - I think ridge is scale-invariant in Y:

$$\hat{\beta}^{ridge} = (X'X + \lambda I)^{-1} X'Y \qquad \text{Now let } Z = YR$$

$$\hat{\beta}^{ridge}(Z) = (X'X + \lambda I)^{-1} X'YR$$

$$= \hat{\beta}^{ridge} \cdot R$$

So just to be like Matlab, demean Y and standardize X.

Wait - why is $\hat{\beta}^{ridge}(Y) = \hat{\beta}^{ridge}(\underbrace{Y - mean(\bar{Y})}_{=: \hat{Y}})$?

$$\hat{\beta}_0^{rescaled}(\text{adding back } \bar{Y}) = \hat{\beta}_0^{rescaled}(\text{not}) + \bar{Y}.$$

0 : intercept

which makes sense. But there shouldn't be this entire business about scaling Y if it doesn't matter for $\hat{\beta}^{ridge}$ ...

Ah ... I think I know $X'Y$ is kinda $Cov(X,Y)$

$$Cov(X,Y) = Cov(X, Y - \bar{Y})$$

↑ b/c $\bar{Y}$ is a constant $(Var(\bar{Y}) = 0)$

$$Cov(X,Y) = E\left[(X - E(X))(Y - E(Y))\right]$$

$$Cov(X, Y - \bar{Y}) = E\left[(X - E(X))(Y - \bar{Y} - E(Y - \bar{Y}))\right]$$

$$= E\left[(X - E(X))(Y - E(Y) \cancel{- \bar{Y} + \bar{Y}})\right] = Cov(X,Y).$$

→ So demeaning Y gives you the <u>same</u> $\hat{\beta}^{ridge}$ as not demeaning.

So it seems like in any case you adjust the intercept, adding $\bar{Y}$ to it. Ok...

Finally the last ridge-thing of interest:

Y $n \times k$ (for VAR)

↳ Matlab can't do it, but doing it by hand, and then separately for each regressor gives the same thing.

↳ So move on: generate data w/ the new ridge commands and see if that solves the VAR-instability issues.

Note: the data-generation itself of 13 August ($b_u = 2$) complains of VAR-instability. That had the old ridge.

Data-generation complains w/ new ridge code?
(rf_var_ridge.m)

Nope, it didn't! :) ($b_u = 2$, $\lambda = 0.001$)

⇒ acf_sim_univariate_data_18_Aug_2020.

Ok, so finally do est w/ 18 Aug data

- ridge $\lambda = 0.001$
- $b_u = 2$
- $\sqrt{W}$ (! this guy needs to become the new reference fig, as in Fig 3, Materials 41)
- $\lambda = 0.001$

Estimation : instability in VAR : 6 times

↳ It seems like it's the initial 6 evals that cause problems. Warning never occurs afterwards.

→ Fig 4, Materials 41 (Flag = 3)

- Try same thing for TolFun = 1e-9 → same, Flag = 2.
- same, TolX = 1e-9. → same, Flag = 3.

- Now: do 'manual' grid out in far end [TolFun = TolX = e-9]

But really one should have TolX = TolFun = 1e-6 (default) and add Wmid = 1000. Yeah this one's not great.

↳ Doing that now.

- tols back at defaults
- $b_u = 2$ (ridge, $\lambda = 0.001$)
- $\sqrt{W}$
- fegrid = [4, -3, 0, 3, 4]

Not bad but not great either. What avenue to pursue? Raise b_u more?

Prep for Peter:

- ZEW interview send link!
- Presentations: Aug 25 12.30 - 2pm (BU macro read)
  Sep 1 3.30 - 4.30 pm (Diss Workshop)
- Content: just Fig 15 from Materials 40

Peter meeting 18 Aug 2020

Main issue: works, but not so well in the neighbor-hood of 0 fe.

↳ Ready to go back to true data? Yes?
But what you wanna do for sure is to constrain the parameters of fe close to zero.

For real data: proceed in steps:

- first identify α (large fe) and constrain α(small fe) and see how well you can do
- gradually expand # of α to expand to α(small fe)

- Volatility of shocks important AND data is likely to be informative about them
  ↳ could est them
  ↳ or calib them such that model-sim data matches volatility of data.

· δ shocks are simply harder to calibrate!

→ 2 questions:

1. In real world data, you wanna have a good feel for how ambitious you can be in est α and whether this depends on δ
2. δ is simply hard to assign values to.

↳ Talk after BU seminar!

10:30 – 11:45 Tu & Thr   Fabio 9 – 10.15

Do an email exchange after DissWork to talk on Sep 2 or the next days.

<u>Work after</u>

Ok – so see if we can estimate δ.

I wanna do a new obj function which complements
obj_GMM_LOMgain_univariate_mean.m _shocks.m
in that it also estimates δ_j $j = r, u, i$

I also wanna
✓. change names of figures so that they're not as long

Try it on Aug 10 data b/c that's the 4x scaled-up truth.

(!) And $\sqrt{W}$ is a default now!   (!) Wmid=1000 too! Nsimul too!

The code is running... Took 713 sec. And it ends: ↪ 70 iter

$\hat{b}_j = 0.1 \quad \forall j$ (while truth is 1).

Why didn't it move? Error in code or? Is it time to try Ryan's thing of making it take larger steps when evaling the loss?

After 70 iterations, it looks like Matlab never tries anything else than b_0. It doesn't seem like it's a problem of getting the values into "param".

→ I think it's something like "forward finite differences"

As usual, some background: finite differences seems to be a numerical approximation method for derivatives.

Wait you idiot: of course it's not trying out any values of b: b is irrelevant in the model as it now stands. I think I need to modify the obj. fct.

19 Aug 2020

→ Yeah, now I can see Matlab trying different things, that's good.

Finite difference := an expression of the form $f(x+b) - f(x+a)$

difference quotient := $\dfrac{f(x+b) - f(x+a)}{b-a}$

Finite differences are used in approximating derivatives, — "finite diff approximation" or for "finite difference methods" for solving ODE's.

finite difference quotients are used in approxing deriatives.

3 types:

forward difference: $\Delta_h[f](x) := f(x+h) - f(x)$

backward difference: $\nabla_h[f](x) := f(x) - f(x-h)$

central difference: $\delta_h[f](x) := f(x+\frac{1}{2}h) - f(x-\frac{1}{2}h)$

! Authors who use the term "finite differences" to mean "finite difference approximation of derivatives" also mean "forward difference quotient" when they say "forward difference".

Relation w/ derivatives

$$f'(x) := \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

For a fixed h, this is

$$\frac{f(x+h) - f(x)}{h} =: \frac{\Delta_h[f](x)}{h}$$

the forward difference quotient!

And I think the 15 July correction to the convexity moment does this: $\Delta\alpha = f(x+h) - f(x)$, $h = \Delta f_{egrid}$.

Exciting stuff: in Matlab's solvers, there's the

FiniteDifferenceStepSize : I guess this is "h".

The default is sqrt(eps) for forward

(for optimset, "FinDiffRelStep") eps^(1/3) for central finite diffs.

FiniteDifferenceType: "forward" (default) or "central"

central is more accurate but takes twice as many function evaluations.

(for optimset, "FinDiffType")

I'm trying: same (i.e. MaxFunEvals = 2100 (instead of 700)) → 45 min!

but FiniteDifferenceStepsize = 100 · sqrt(eps).

Same. $\hat{b} = (0.43, 0.29, 0.56)$.

Let's go back to MaxFunEvals = 700 (12 min).

and set FinDiffStepSize = 100 000 · sqrt(eps). awful! timed out & didn't move

↳ Better but still timed out. Try 1M · sqrt(eps)

Let's think a little: why is the solver iterating out now w/o convergence? Am I starting too far from the truth? Is this complementarity between α and b which comes from the ⊕ feedback loop in the model?

What is really weird is that, compared to the case where

I fixed $\gamma = 1$, now the estimates for $\hat{\alpha}$ are <u>lower</u>, while $\hat{\gamma} < \gamma^{true} = 1$. I would have expected $\hat{\alpha}$ to be <u>higher</u>, b/c the true data is more volatile (b/c $\gamma^{true} > \hat{\gamma}$) so to capture that, if the estimation doesn't raise γ, it should raise $\hat{\alpha}$! Analogously, the moments aren't matched well at all.

Interestingly, increasing the findiff stepsize does allow the estimation to move away from the initial values more and fits moments better too. But why doesn't the solver continue on this path?

With findiff stepsize $= 100k \cdot \sqrt{\varepsilon}$, $\hat{\gamma} = (3.5, 0.9, 0.75)$ and $\hat{\alpha} < \alpha^{true}$, but closer. So, in a sense, I see the tradeoff I was talking about (instead of $\hat{\alpha}\uparrow$, $\hat{\gamma}\uparrow$, in fact above γ^{true}), but in a way, $\hat{\gamma}\uparrow$ was accompanied by $\hat{\alpha}\uparrow$, so the complementarity is kinda also not there.

I also wonder whether there's a kinda "symmetry indifference" in the sense that in half of simulations, $fe < 0$, in the other half, $fe > 0$, so $\alpha_{1,2} \gg 0$, $\alpha_{4,5} = 0$ matches moments just as well as $\alpha_{1,2} = 0$, $\alpha_{4,5} \gg 0$. Similarly, $\gamma_r = 0$ & $\gamma_i \gg 0$ may be undistinguishable from $\gamma_r \gg 0$, $\gamma_i = 0$.

Or is the estimation saying that x is more volatile and therefore $\hat{z}_r > \hat{z}_i$ or $\hat{z}_u$? But why isn't it raising $\hat{\alpha}$?

And is 1M · default findiff stepsize not moving at all b/c

i) $\hat{\text{Jacobian}}$ is so imprecise

ii) or the loss is really flat over large surfaces?

$$\frac{f(x+h) - f(x)}{h}$$ is initially < 0 as $\hat{z}$ is raised,

and then it becomes consistently $= \varepsilon < 0$, small.

Maybe what is going on, is that if $h :=$ findiff stepsize is small, it's not getting to large z, so it times out, but if h large, it gets to large z, but then the steps in α are too large to make a diff.

Let's try $h = \begin{bmatrix} h_z \\ h_\alpha \end{bmatrix} = \begin{bmatrix} 100k \text{ for all the } z\text{'s} \\ 1 \text{ for all the } \alpha\text{'s} \end{bmatrix}$.

Interesting: this leads to $\hat{\alpha} < \hat{\alpha}$ ($h = 100k \cdot$ default $\forall$ params)

$\hat{z} < \hat{z}$ (—11—)

Reverse to $h = \begin{bmatrix} h_z \\ h_\alpha \end{bmatrix} = \begin{bmatrix} 1 \text{ for all } z\text{'s} \\ 100k \text{ for all } \alpha\text{'s} \end{bmatrix}$

Interesting: this gets me almost as far as $h = 100k$ for all!

↳ So it's the h in α that matters, not so much the one in z.

What does that mean? → The loss is locally flatter in α than in z.

Try: $h = 100K \times$ default for all (uniformly), but initialize at truth: $z_0 = 1$.

↳ Much better (see Materials 41) → seems to suggest that some of the "small step issue" may be coming from the initialization. But maybe the same holds true for α?

↳ Trying $z_0 = 2$. Much more explosions and $k^{-1} < 0$. But best so far! If I don't make the estimation "experience volatility", it doesn't.

↳ Maybe should do the same for α: initialize at truth or at higher values.

Try $10\,\alpha_0 \approx (0.176, 0.099, 0, 0.099, 0.176)$ → doesn't work (VAR) $5 \cdot \alpha_0$ doesn't either.

Also: I'm wondering if I should penalize $k^{-1} < 0$ in the estimation?

Prep for Ryan

- ZEW interview
- Presentations: Aug 25 → not avail. 12.30 - 2pm (BU macro read)
  Sep 1 3.30 - 4.30 pm (Diss Workshop)
- Teaching: coordinate → no rules or things to abide by.

Ryan meeting 19 August 2020

Makes sense to $z_0 > z^{true}$.

- Pick τ_j such that you hit moments of data.

- Fixed-point problem: write a truth that looks like the data.
  1) fix some params. → Est. on real data. Take $\hat{\alpha}^{real}$ and put it into the fake data est. → Iterate until convergence.

Loss is
- Not continuous — capital crime
- Kinky — misdemeanor

} 2 crimes you can make

↳ cutting simulation that exploded introduces a jump in the obj. fct.

Could explain diff to NEshin

↳ Any gain that's <0, set it to zero and go on w/ the simulation

↳ for exploding fc, you could set fc to a max

↳ clean sol.

vs projection facility which introduces kinks!

Worst-case scenario:
Can take 3×3 sigmas, est α for 9 cases, take avg$(\hat{\alpha})$ or the one w/ the best loss.

Try to calibrate τ so as to match moments of data. See if est then is better ID-ed. Can send him an email to get feedback.

Take-aways:

① 1^{st} priority is to calibrate τ's to match data moments (for a given α) and see if this higher volatility world helps ID-ing the α's.

② Need to address how you deal w/ exploding fe and $k^{-1} < 0$ in individual simulations. Throwing those out introduces jumps in the objective fct which is the biggest crime you can commit.

↳ doing this could explain the difference to the "N estimations" strategy b/c that one doesn't have these jumps, it has a cross-section of $N = 200$ anyway!

③ For the "ceteris paribus problem" in τ and α, can start w/ a laboratory w/ fixed α, τ. Estimate those for fake data. Does it work? Estimate on real data. Use $\hat{\alpha}$ & $\hat{\tau}$ as α^{true} & τ^{true} in the next round of fake data. Iterate.

The first priority is actually to deal w/ exploding fe and $k^{-1} < 0$.

20 Aug 2020

↳ set to 0.

↳ set to $|fe| < 20$?

First of all: fe = nan & $k^{-1} < 0$ don't occur together a lot.
But displaying the $k^{-1} < 0$, I see that it's clearly $\to -\infty$. Isn't fe exploding then too? Yes it is.
Interesting that it's not being caught by the test isnan(fe)
It seems like if the simulations would be longer, c+209 might just turn into nan.

↳ If, within the simulation, I catch $|fe| >$ threshold

$|k^{-1}| >$ threshold

$k^{-1} < 0$

then I get in control of the situ.

Set $k^{-1} < 0$ to $k = 10^{16} \to k^{-1} = 10^{-16}$.
Her 16 is where the first $k^{-1} < 0$ occurs.

↳ The strategy seems to work! Didn't change estim tho.

Now turning to the exploding simulation case.
- It does seem to begin w/ large fe and gain

Try to set a tight bound on fe & gain, for now not on the observations.

What's the appropriate threshold?

$|fe| > 10$ already seems to have $k^{-1} > 0.6$

And it seems like once $|fe| > 10$, you always explode eventually.

Ok, let's just try $\overline{fe} = 10$, $\overline{k^{-1}} = 0.6$.

As a first step, let's cap fe at $\overline{fe} = 10$, and do nothing else.

Did impact fe & estimation. It never explodes now

is worse

but there are VAR instability issues which suggests that either observables explode or they are simply constant. → Yeah b/c now things that are exploding enter the VAR which didn't before. I can also see that observables, gain & fe continue exploding. So once fe reaches 10, the economy is already on an explosive path.

Since here $\alpha^{true} = [0.2, 0.1, 0; 0.1; 0.2]$ for $fe = [-2,-1,0,1,2]$ it follows that for $fe = 10$, $k^{-1} = 1$. So $\overline{fe} \overset{!}{<} 10$.

Try $\overline{fe} = 5$.

- First of all, now no VAR complaints.
- Second of all, obs & gain don't seem to explode now.
  But I've found some explosive paths, they just don't have time to explode.

Loss(α_0), $\overline{fe} = 5$ → no complaints.

Obj(α^{true}):

One simulation ($n = 87$) explodes. What are its features?

1) $fe > 5$ at $t = 111$. $k^{-1} = 0.1712$ et basta.

But I wonder if we could penalize k^{-1} instead b/c the threshold fe is likely to depend on α, while k^{-1} may not.

Now I tried to do it with restricting $k < \overline{fe}$. It seems to be stricter, 3 times it's invoked.

↳ And in estimation?

seems more strict here too. Smaller fe's lead to involation too. Smaller obs.

An issue to also keep in mind is that ndim-simplex(·) shouldn't take nan-values, so it's desirable to work on the fe instead of the gain.

How many iter does it take? 73 and now 562 sec. (?!)

Too many outputs! I've reverted to outputting only the % sim w/ explosion or $k^{-1} < 0$, and if within a simul, more explosions occur, I output the % of sim in which more than 1 explosion occured. I hope that can be a metric for whether my fe-threshold is too high.

Way more $\bar{k}^{-1} < 0$ seems to happen. But also more explosions. I don't get it.

↳ That makes sense b/c the crit. for explosions is way lower.

What is also worrying is that $\bar{k}^{-1} < 0$ even if α is convex.

When α is nonconvex, that's when you get the "$\bar{k}^{-1}$ was negative on a fine grid" message.

21 Aug 2020

↳ Is setting the obj fct to NaN here a good idea?

The other $\bar{k}^{-1} < 0$ sitns don't occur for nonconvex α.

$\bar{fe} = 5$. explode_t lets me look into each history at see where the fe-threshold was evoked. I do this for $\hat{\alpha}$. It's evoked in 17% of histories. Sometimes it's evoked at one t only, sometimes for long streaks of time. Interesting is also that even some longer streaks can end e.g. a streak from $t=22$ to $t=50$.

negk_t $= 0 \; \forall t, n$. Hmm. So maybe my counter is screwed up? Yep, the thing was displaying as $\bar{k}^{-1} < 0$ sitns where explosions happened more than once.

$\bar{fe} = 10$ Only 2 iter complain of VAR issues.

At $loss(\hat{\alpha})$, $\bar{fe}$, $\bar{k}^{-1}$ don't seem to be invoked.

$\bar{fc} = 8$

of course gets invoked a lot more than $\bar{fc} = 10$.
and the estimate isn't so good either!
But it stopped a lot earlier (263 sec, 39 iter)

$\bar{fc} = 3$

flag = 2 and estimation is quite off. Grrr!
It took 730 sec and the fc was invoked a lot of course, about 37% of cross-section histories evoke it, 9% several times.
All iter. No $k^{-1} < 0$.
explode_t shows that crossings are only punctual, there aren't really crossing spells.

Now turning to see whether b's can generate some models w/ moments close to those of the data:
command_sigmas.m

Peter meeting — 21 August 2020

- b's are scaling the red lines in Materials 42
- If the underlying NK can't come to grips w/ the volatility of the most important observables of the model, then need to address that first.

- Scaling: if $i = 3\%$ in data, is it 3 or 0.03 in the model.

units: 0.1 may be huge for output gap shock (z_r)
↳ 0.001

The other thing:
- go to real data
- calibrate z_j so as to match the moments of the data

π: multiply model $\cdot 4$, or $\frac{1}{4} \cdot$ data

→ Proceed iteratively:
try: 2, 2, 0.5 → as long as π isn't volatile enough, raise z_u
But supp. you try this and there's no way to get it anywhere near, then the NK core parameters may need to be recalibrated.

The last thing:
NK model core: b, κ, ψ_x gleaned from data.
Calib shock z to match vol(π, x, i) in data.
Question: are there any params left? No.
Suppose you impose: 0 at 0 → reduce to 2
symmetric → parameters
convex

trying to calibrate the remaining two α's to match the vol(Expected (π)) in the data & the corr. btwn Exp(π) & π_{t-1}.

↳ b/c α's affect : $E(\pi)$ & how π responds to $E(\pi)$ (or vice versa, better I think)

2 advantages to proceeding this way:

1) indication that you can go back to real data & do a full-blown est. (if this works)

2) could use this careful calibration as a basis for policy exercise.

No meeting next week.

Touch base mid-next week via email so that if I need to meet before the Sep 7 seminar we can still meet.

- Many jobs come in the Sep Ed. of JOE (mid-Sept)
- But many jobs miss that deadline (⇒ so mid-Oct)

most: Sept 15 - Oct 30

deadlines: Nov 15 – Nov 30

Work after 23 Aug 2020

- Go back to get_data.m
- The model is quarterly so π should be quarterly (q-o-q) NOT y-o-y. (%) → corrected. ✓
- Checked that x and i are both quarterly and % ✓
- Now the only thing I'm stuck at is $E_t(\pi_{t+1})$. In the model, this should be quarterly: $E_{Q3}(\pi_{Q4})$ e.g.

But the data is 1-yr-ahead, e.g. $E_{2000-Q3}(\pi_{2001-Q3})$. Also, and more importantly, the data thinks of annual (or annualized) rates, i.e. 2% annual rate in Q4. In the model, the rate is quarterly.

→ This can be solved by taking "CPI3" from the "Mean_CPI_Level.xlsx" from the SPF.

1 = previous quarter
2 = current quarter
3 = next quarter
etc
A = this year
B = next year

But I will still need to annualize the quarterly model rates.
↳ Or actually, if π is quarterly, I'd need to deannualize the data!

.

compromise: what if we take quarterly frequency π & $E(\pi)$, but an annualized rate? Easier to interpret.

Wait: · the rate should be annual (% inflation per year)

· the frequency quarterly (occurs every quarter)

· q-o-q ?

$\Rightarrow$ 1-quarter ahead, annual rate

"In the next quarter, inflation on an annual basis will be 2%"

$\leftarrow$ quarterly change in %.

In the model, $\pi_t = \dfrac{P_t - P_{t-1}}{P_{t-1}}$ $\Rightarrow$ q-o-q

and in each quarter, an obs. is generated.

If you annualize this, you <u>don't</u> get the y-o-y change!

Ok so now I have q-o-q changes in % in π in data and model, but for expectations I have q-o-q $E_t(\pi_{t+1})$ but at annualized rate.

$\hookrightarrow$ So I can either annualize model π, $E(\pi)$ & data π, or de-annualize data $E(\pi)$.

SPF: documentation: "All growth rates are expressed in annualized percentage points. The formulas use discrete compounding, not continuous compounding."

Annual growth rate for CPI inflation in SPF $\left(\frac{4^{th}\text{ quarter}}{4^{th}\text{ quarter}}\text{ grt}\right)$

$Z_t :=$ 4^{th}-quarter value of a variable in year t.

$gZ_t :=$ annual growth rate of the variable:

$$gZ_t = 100 \cdot \left[\left(\frac{Z_t}{Z_{t-1}}\right) - 1\right]$$

Discrete compounding

$$A = P\left(1 + \frac{r}{n}\right)^{nt} \rightarrow A = P\left(1 + \frac{r}{4}\right)^{4}$$

A = amount after time t

P = principal (initial amount)

r = interest rate (annual)

n = # periods in a year (compounding frequency)

t = time in years

Continuous compounding: $A = Pe^{rt}$

I have data: $\pi^d \rightarrow$ this can be q-o-q or y-o-y, annual rate quarterly.

and the SPF $E(\pi)^d \rightarrow$ annualized q-o-q % changes

And model: π^m is q-o-q % change

$E(\pi)^m$ is q-o-q % change

$$A = P\left(1+\frac{r}{n}\right)^{nt}$$

So for $E(\pi)^d$: $\quad CPI_{Q2} = CPI_{Q1}\left(1+\frac{\pi^{annual}}{4}\right)^4$

$$\Rightarrow \frac{CPI_{Q2}}{CPI_{Q1}} = \left(1+\frac{\pi^a}{4}\right)^4$$

$$\Rightarrow \left(\sqrt[4]{\frac{CPI_{Q2}}{CPI_{Q1}}} - 1\right)4 = \pi^a$$

What I have in the data is $\pi^a = E(\pi)^d$. Can I convert this to q-o-q changes, that is, $\frac{CPI_{Q2} - CPI_{Q1}}{CPI_{Q1}}$?

I think yes, as

$$\underbrace{\frac{CPI_{Q2}}{CPI_{Q1}} - 1}_{E(\pi)^{q\text{-}o\text{-}q}} = \left(1+\frac{\pi^a}{4}\right)^4 - 1$$

But this also means that if I have $E(\pi)^{q\text{-}o\text{-}q}$ or $\pi^{q\text{-}o\text{-}q}$, and I want the annualized rate π^a, I write

$$\pi^{qoq} = \left(1+\frac{\pi^a}{4}\right)^4 - 1 \Leftrightarrow \left(\sqrt[4]{(\pi^{qoq}+1)} - 1\right)4$$

$$\Rightarrow \pi^a = 4\left(\sqrt[4]{(\pi^{qoq}+1)} - 1\right).$$

Annualized π-rate from q-o-q π.

Dallas Fed on Annualizing Data

$g_m :=$ annualized % change.

$X_m :=$ values of variable in month m (level)

$$g_m := \left[\left(\frac{X_m}{X_{m-1}}\right)^{12} - 1\right] \cdot 100$$

For quarterly data:

$$g_q := \left[\left(\frac{X_q}{X_{q-1}}\right)^{4} - 1\right] \cdot 100$$

Suppose, as in their Table 1, the July numbers for employment are

	(thousands)	Monthly % change	monthly % change annualized
June	9553.8	$\frac{9574.8 - 9553.8}{9553.8} \cdot 100 = 0.22\%$	$\left[\left(\frac{9574.8}{9553.8}\right)^{12} - 1\right] \cdot 100 = 2.67\%$
July	9574.8		

$$g_q = \left[\left(\text{q-o-q grt} + 1\right)^{4} - 1\right] \cdot 100 .$$

Annualization := Adjusting a growth rate to reflect the amount a variable would have changed over a year's time had it continued to grow at the given rate.

So if the SPF is an annualized qoq growth rate, then I should use the qoq π^d, and for π^m, $E(\pi)^m$ and π^d, I should annualize as

$$\pi^a = \left[\left(\frac{\pi^q}{100}+1\right)^4 - 1\right]\cdot 100 .$$

Annualized inflation (expectation), net, in %, is the gross quarterly inflation (expectation) taken to the 4^{th} power, subtracting 1 to get the net, and multiplying by 100 to get the %.

I'm in command_sigmas.m and I notice that η is wrong. It should be $\eta = \begin{bmatrix} b_r & & 0 \\ & b_u & \\ 0 & & b_i \end{bmatrix}$ and not $\begin{bmatrix} b_r & & \\ & b_i & \\ & & b_u \end{bmatrix}$

No, sorry, that's actually not true b/c the order of shocks is $s_t = \begin{bmatrix} r_t^n \\ i_t \\ u_t \end{bmatrix}$, so $\eta = \begin{bmatrix} b_r & & \\ & b_i & \\ & & b_u \end{bmatrix}$!

Ok so redid PEA & VFI w/ the calibration. <u>To DO tomorrow:</u> figure out why objective_CB_approx.m is giving trouble!

- learning_code issue ← plot_sim_loss.m
- model_NK issue ← grid_search_approx.m

For grid-search.m, when I use the correct learning code, the solver wants to try $\psi_\pi = \text{NaN}$, for which reason model_NK.m returns an error, b/c it cannot take the derivative of a NaN. → But why try a NaN???

As for plot_sim_loss.m, the errors occur in ndim_simplex.m b/c the point at which it's evaluated, the current fc, is NaN.

↳ My explode_count variable reveals here that indeed as you raise $\psi_\pi \geq 1.7$, you begin getting explosions. It seems you need to keep $\psi_\pi \leq 1.4$ not to explode.

↳ Yes I've checked and above that the loss is only zero b/c bet^t × (inf + inf + inf) = NaN.

↳ to continue w/ why the solver tries a Nan!

It looks like grid-search.m works if I just set ub = 1.4 (given that I know the loss explodes for a $\psi_\pi > 1.4$). <u>24 Aug 2020</u>

↳ It's worth it to always do the plot_sim_loss.m first so you have an idea how the loss behaves and what reasonable bounds are.

BU prezi notes 25 Aug 2020

Adam: What about an NK model w/ an inflation grt (which is a random variable). For some reason, people have a dispersed prior.

Adam: Crazy interest rates – doesn't seem feasible. I'm also wondering whether a first-order approx is fine when int. rates move 20% – 30%.

Slide 25: ZLB will get CB in trouble. Also would CB want to get $E(\cdot)\uparrow$ as much under ZLB too? Slide 25 will get you in trouble b/c responses so large.

After prezi:

Stephen: Cool & carefully done. 2 categories of comments.

① Presentation: difficult to show empirical results w/o model. Instead: Show a data fact! E.g. "I regress $E(\bar{\pi})$ on fc and I get"

② Model: 1) Why just constant & not the slope?

2) Agents estimate gain from data, but not in an optimal way.

3) Comment on E-stability.

4) Do you care about adaptive learning or only about aggressive policy? → Those two imply different approaches.

Adam:

1) Most delicate point of prezi: when you introduce the simplifications: why don't they learn the slope? Once you get ppl on board w/ that, everything follows. Need to convince them though so can pass over it smoothly. Sthg like: "Given I want to model anchoring, I'm going to do the smallest possible deviation (from RE)"

2) Worried by how aggressive policy is → will lose ppl. In particular, that a hiring committee member will not have understood anything, just remembers that the numbers were crazy and so wouldn't hire me.

Is there a way to do the simple version for analytical purposes, and then the full-blown version for the quantitative part?

Ryan meeting 26 Aug 2020

slide 25 Stdder of CBO output gap? x fluctuating more by an order of magnitude.

Check whether int. rate is annualized too.

↳ Early 80s did have a lot of interest craziness. they might drive the changes in ACF.

- Cross-covariances targeted is pretty ambitious when only 3 params
- Why is $\mathrm{cov}(x, i)$ not < 0 in the data? Check the series?
- Trend is gonna dominate the variance term (for $\bar{\pi}$)

  $\bar{\pi}$ is 4 times as volatile as $E(\cdot)$. May not be wrong.
- Calibration B is a good backup and good for the exercise to see whether the model can estimate α now.

Interest rate (DFF) is yield per annum, 27 Aug. 2020

that is, the damned thing is annualized goor!

↳ Need to annualize the model!

~~ok - but I'm gonna agree w/ myself that for the opt. TR exercise and the PEA & VFI I don't need to annualize b/c for PEA it just scales ego up or down, it doesn't change when they hold.~~

No, I think I have to annualize these too b/c the concern is that if there are nonlinearities, then that changes the shapes.

The problem though is that for this calib, the loss in plot_sim_loss.m always explodes. So I'm NOT gonna annualize PEA & VFI and TR.

Fck this! Now the PEA code doesn't work - it keeps running forever, seemingly w/o progressing. The issue seems to be the parallel pool...

↳ Yes, that was it! Somehow the parallel pool stopped working! Restarted and now it works!

Loss: Anchoring / RE

$\psi_\pi = 1.5$, $\psi_x = 0.3$ 1.5559

RE-optimal 2.5660

Anchoring-opt 1.1199

Sum: 5.2418 | :3

Avg 1.747

⚠ Theoretically, I would need to annualize π, i, $\bar{\pi}$ & f_c in PEA. But I agreed w/ myself NOT to right now and therefore I'm also not annualizing them on the plots b/c they wouldn't correspond!

Stephen's data idea: What I can do is: 28 Aug 2020

- construct f_c in the data
- reg long-run $E(\pi)$ on f_c

To do: need to figure why there are 29 Aug 2020
matrix instability issues when I use expectations in estimation w/ calibration C.

↳ Need to implement Ryan's rescaling of W b/c ratio of smallest to largest element is bigger than it should be.

Implemented "Ryan's rescaling" but it doesn't solve the instability issue. What does though is $\uparrow \lambda^{ridge} = 0.01$ (from 0.001)

31 Aug 2020

Dissertation workshop prezi

1 Sept 2020

Ryan R: stabilization vs. volatility tradeoff

Alexey A: inflation-responsiveness

R: Eq. (8)

Susanto S: deviate from adaptive learning

Jenny: anticipated volatility: in some cases it does matter b/c subjective uncertainty matters for precautionary savings

Jenny: constant gain is a benchmark

S: Is the reg well specified? Is the error term uncorr w/ regressor?

J.

D: individual data can be useful to estimate

(Daniel) the anchoring fct.

Ryan comments

Fit & finish & marketing

✓ • Powell's comment not so much related to my stuff

☑ S. 3 is not consistent w/ the SPF.

↳ Use the SPF series to remain consistent

✓ S. 4. Good: Remove "optimal" from 1st line

→ narrows audience

Frame your learning as "extension" "not special case"

✓ S.5 "2nd ↳" feels clear but the first isn't

↳ don't put it in

b/c it's not crystallized & doesn't give the casual person the feeling they're understood.

↳ may want to put in later as it's clear

✓ S.9. Susanto's Q: "10-yr as LR and therefore it doesn't make sense in RE: b/c var = 0 so cov = 0.

But if you interpret it as just a longer-horizon π-Exp, but not the LR. His q was: "Is the reg well specified?" But I'm not saying that it's a structural relation I've estimated. ⟶

✓ Put reg on page. Is it a rejection of RE? Even if it's just a longer-term Π-Exp(·).

Core is less volatile. Might plot that.

✓ S.14 S "what's in the head of ppl?"
Distinguish btwn what ppl think about the future (anticipated utility) vs. the reason ppl adopt this rule (Peter)
"This is a behavioral description of what agents do not of what they think" → could answer by saying that what they do makes sense b/c it fits the data. The informal description is that they are worried about the regime changing
↳ practise this convo b/c source of quicksand.

✓ Eq. (11) $\sum \alpha_i b_i$ may be easier!

✓ Wanna call target crit. a Proposition?

✓ S.24 talk only about PEA & cubic spline

S.25 onwards: How present results:

☐ · Working w/ a couple different calibs in my head and the audience can't bounce around.

✓ · Should explain calibration
Split table in 2 : standard
picked to match what exactly

Whether you show moments depends on how you do in the data → for calib, if you do well, you don't have

S. 26 ☐ Add axis labels!

✓ $i(\bar{\pi}$, everything else at its mean$)$, not $\partial i/\partial\bar{\pi}$

☐ Show this line for some low k_+, and some high k_+ (90th percentile) ↳ 10th percentile

Also this line for (gain learning would be interesting but might be too much info. (for a high & low k too)

☐

S.27 k_+ in 10th vs. k_+ 90th percentile

☐ S. 29 Axes!

☑ Could do the loss for $\lambda_x = 0.05$ to avoid conclusion

☐ Line for optimal loss!

Susanto: Taylor-rule w/ $E(\bar{a})$ should do it

↳ should mention that you did.

✓ S.31 "I can get 90% of the distance from 13.6 to 5.31 by acknowledging 6.14 anchoring"

Take the RE in RE ✓✗ Anch in Anch. ✗ RE in Anch. ✗

S: Is this relevant to the real world?

Practice answering this succinctly w/ one ex. under several headings.

1.) Yes it's a plausible learning model which makes sense in terms of what we do (introspection)

2.) And name 1 paper that shows it empirically, especially if it's (experimental) time-varying gain

3.) The pattern of E(·) in the data is consistent w/ the model. Coib & Gorod of the reg (forecast-level)

✓

4) These models do a good job of matching the data, cite E&P & Milani. (time-series data)

Point 2) Cars Hommes type est. learning models on people in the lab

introspection + experiments → micro-founded expectations β + outcomes β → macro-relevant

↳ Play around w/ calibration to see if can decrease i-volatility.

<u>Work after</u> To think about:

(1) In motivating regression, is $\hat{\beta} > 0$ a rejection of RE? Is it a rejection of RE even if SPF10 isn't a long-run expectation, just an expectation far in the future?

↳ I think no, b/c $E_t(\pi_{t+k})$ can and does move in RE. I think though that for a large enough k, the movement should be minimal b/c if agents know the model, then $E_t(y_{t+k}) = g \cdot E_t(s_{t+k})$ and in $h<1$, $h^k \to 0$.

$-x^2 - (x+1)^2$ PS 1

$-x^2 - x^2 - 2x - 1$

$= -2x^2 - 2x - 1$

FOC: $-4x - 2 = 0$

$-4x = 2$

$x = -\frac{1}{2}$ $y = \frac{1}{2}$

Damial meeting 3 Sept 2020

0) Key parts are these → think hard of how to communicate

1) Intro: language assumes audience like Ryan lingo, and even Ryan didn't understand
But is intuitive enough.
Powell quote is good to make the case
but read it less technically, break it down
E.g. start w/ RE-benchmark & explain what ~~explain~~ anchoring is. In 30 sec (initially 5 min)
↳ guide us to what def you'll use
"If we were in RE, there's no way ..."
Focus on first 10 min, first few slides!

Slide 2) Little fig to anchoring!
Why should I be surprised

→ LR-exp = st. st in RE. Policymakers are worried that unanchored: this doesn't hold
→ Anchoring def.

Broad comment #1
2) Easy to be disengaged in Zoom: bring enthusiasm & energy. Very low energy in talk! Come up w/ a routine to be excited.

3) Broad comment #2: Answers too long in general. Very conscious about it. Ok to be quiet for 5-10 sec so that you can answer very tightly & succinctly! Avoid repetitions, digressions. Don't tell everyone that it's a good question.

Slide 4: · Not happy b/c didn't know what anchoring

- [ ] · Should avoid "endogen. gain learning"
  ↳ could give intuition of it though (don't cite papers though)

· Don't know what anchoring yet is

Slide 5 even worse

- [ ] · Need to be reminded of optimal policy But not patronizing. Make them feel that it's simple
  E.g. respond to <u>what</u>
  If too much to say, then don't mention it.

→ Think of what is simple to communicate as the main result

☐ aggressive & length → don't know what it means

Every economist should be able to tell in 5 min what is going on.

How to impress Ryan:
not by technical language, but by how you can translate that language to terms everyone can understand.

Slide 9. Have to be VERY clear about reg.
→ why should I be surprised

Model setup:
- lots of assumptions not fully spelled out.
  ↳ specify what's in their info-set

☐ Slide 13 → spell out very clearly what the assumptions on E are
What are the set of beliefs that lead to this updating rule

These agents seem really dumb, in his view they aren't as dumb.

- they believe the world is stationary whereas the world is not stationary
  → very dumb & doesn't make sense

a) When Susanto asks you why it's a good model

- your strongest point has to come first
  ↳ "intuitive" is your <u>last</u> resort

  empirical support is much stronger
  ↳ start there

  1. RE benchmark fails
  2. This rule captures lot of E in data
  3. All policymakers are concerned about.

Peter meeting   3 Sept 2020

1) Key elements were there
2) But in a job seminar what you want is that ppl even in micro theory can walk out w/ the message "the RE ass regarding the Fed's target of 2% implies it will be 2% and yet in data we see E not only ranging away from 2%, but ranging away from it in a way that's systematically related to fc and what we learned is that it's not enough for monpol to tell the PS, but it ..."

Specific comments:
1) Most was there, but repeatedly what happened is that b/c some details we left out, ppl asked questions which took longer.

E.g. the regressions

↳ ppl were thinking I'm testing a theory or a model but what I'm saying is "if the 2% target is fully credible, LR-Eπ should be 2"

↳ the regression shows sensitivity to fc

"E deviate all the time from 2% and they do so in response to observed movements in π"

↳ so if you're a CB-er, what do you need to do w/ the tools you have to prevent unanch.

But for this we cannot rely on reg. b/c only corr. So I develop a fully developed model w/ a departure from RE.

↳ same kinds comments for IRFs w/ TR-coefficients. Also there you wanted to not show the details but then people asked → e.g. the calibration.

↳ Need to explain even std parameter values.

One thing that would help:

when you introduce

$$\bar{\pi}_t = \bar{\pi}_{t-1} + k_T (f_e)$$

"Helpful to consider two cases usually considered"

dgain → optimal learning about a stationary environment where agents take an average of the sample, which converges

cgain → agents are open to the possibility of regime shifts → rolling subsamples

→ "I use a different model of learning more suited to model anchoring endogenously"

S: Why is this a well-suited model?

We observe in data (graph & reg) that CB perceive that seem to reflect fundamental deviations from RE and CB-ers ask 'what should I do'?
I take a canonical model & tailored it to
a) explain the phenomenon
b) tell CB-ers what to do
My paper is the first to answer this questions.

☐ · state-space sol of RE & then learning
Should I show IS- & PC before showing $\hat{E}$
"I take a std NK model which has a state-space sol, and then you introduce $\hat{E}$ and IS & PC."
↳ Maybe show IS & PC before state-space.
· You may notice Π_1 & Π_2 are different
· Bozovic's point, I'm happy to discuss but it's not contrib. of my paper.

"In the paper I analytically characterize opt policy which yields results about the nature of mon. policy. It can be more instructive to use numerical results to tease out features of monpol which are less model-specific & more robust to calibration"

Email Tues & Thurs on Schedule

↳ One could make slides in the appendix, and if have time, you can cover those in the end.

→ went to grad school w/ Bob.

- Marvin Goodfriend: Inflation Scare Problem

Marvin & Bob: "Incredible volcker disinflation" ↳ dynamics can't be explained by RE ⇒ must be that Fed wasn't credible ↳ until the Fed took action costly in SR.

Thought Bob was gonna raise that isn't this the opposite of the inflation scare problem?

↳ Yes b/c E can deanchor (inflation scare) today we're seeing the opposite

- Sargent, Conquest → PS is RE, the CB learns

↳ question was not so much about adaptive vs RE. It was: can Great Infl. be explained if the CB believes there's an exploitable tradeoff when there was none. The challenge to a. learning is how quickly the Fed would have caught on that there's no tradeoff.

Susantu meeting

3 Sept 2020

For job market:

1. what are your expectations:

   regarding my communication w/ you

   a) do you want regular updates on job list

   b) do you want to discuss my applications

   c) want to give you early notice, but this may not always be possible

2. Would normally Lucy (Gail) handle all reference letters? Even if they're not via JOE? (As will be the case a lot this year.)

3. My preferences:

   1.) Research institution in the US / Europe, but flexible on location (can be global)

   2) Central banks, especially Fed, ECB, Canada, Scandinavian CB's, BoE.

   3) Teaching institutions / World Bank, IMF

   4) Research / policy NGO's like RAND, Mathematica ...

   5) Industry: prefer "research teams" at Microsoft to consulting

④ Guidance on what "tier" institutions I can/should target.

↳ unwilling to write references...

⑤ Our communication about my job market materials

- JMP
- application materials

⇒ What is useful for you to have, and what will you comment on?

---

1. He doesn't discuss positions to apply

☐ · Share a Google spreadsheet → have it online & shared.

· Select 12 places you really think are a good fit

· Apply to places where you see any reasonable prob.

· Also on the high end.

2. Even in the past, fair amount of places who don't use JOE & EJM

☐ 3. Follow up before the first application b/c otherwise it's the same letter used.

4. Typically IMF mid-October is the first

5. He'll look at JMP — late September when he writes the letter.

6. Preferences: More postdocs. A good postdoc is a very good for getting a tenure track.

- Up until end Oct - mid Nov: work really hard on JMP
- After that, work really hard on interview & prezi

General feedback

1) Talk is a good macro seminar but not a good general audience talk
   → non-macro audiences
   → introduction in particular (20 min)
   - Why interesting?
   - What you do?

   get a lot of questions
   At some point ppl get that you need to go into details which they won't fully grasp

↳ most important point by far

2) You should talk a bit about
- the fact of anchoring
- why anchoring can't be examined in RE or in other frameworks than yours

First help ppl think in general terms about anchoring → then the std model & why it can't be used to analyze

This is the 2nd most important

3) Communicate lots of energy

4) learning

truth
how truth evolves over time
how ppl think the truth is & evolves

↳ at times it wasn't clear what I'm describing

5) What would happen if people were learning about the slope → that would close off questions

What may distort ppl more is what this corresponds to: they know b but not a

Plausible: ppl know the objective eqs but not mon. target. (But a but not b is the opposite)

→ higher-order uncertainty is behind the world we're entertaining.

W/o lot of time expenditure get credit for doing things
I had 8 minutes left over → so fine.

Send email w/ draft

- [ ] If estim stands before Sept 15 → send email w/ estim results saying that he should use this version of the paper for the Sept 30 letters.

To do

4 Sept 2020

- ☐ Label obs in PEA for opt & TR (quarters) → x axis
- ☐ Label autocorrelogram (lags) → x-axis

Log-dev from st.st on y-axis

- ☑ IRFs in the continuous cost model for anch & unanch

- ☐ optimal policy as a fct of k

- ☑ Add cgain vs dgain in prezi where Peter suggested.

→ ☐ estimation! figure out what's up w/ E

Jenny meeting

4 Sept 2020

☐ · Connect (intro) more w/ other stuff along these lines

1. Move away from full info (Coibion & Goro)
2. Schleifer – diagnostic exp. (noisy info)
   Gabaix – bounded rat.
   ↳ allude to these

Can maybe talk about it when you talk about cgain learning

- What is the key difference btwn these E(·) and why?
- Also for opti policy

↳ No longer sufficient to say that RE doesn't fit, but also why

these other non-FI or non-RE models
- can't fit the data
- drives optimal policy

↳ To do that, you would need to exploit indi-level variation

diagn. E violate some basic rule → deviates from Expected U
and even put E(var) = var (!)
↳ Ref: Macro: "Overreaction in macro"
Fin: "Diagn. E in credit cycles"
↳ oscillating dynamics which they call credit cycles
↳ published

Noisy info / Rat. inattention

4 bp is too small – maybe it's reasonable in a RE / noisy info, especially given that π is persistent.
↳ maybe in other periods it deviates more & I'm sure info on indi level
SPF might be problematic b/c they might

run a VAR which is log diff stationary

☐ ↳ NY Fed consumer survey
- better than UMich b/c larger sample
- b/c stay in the survey longer
- UMich cross-sectionally smooth the data ... impose a normal on it and round it & they're unclear about it.

↳ firm surveys
1. Livingston : has LR-E
2. Yuriy have a firm-level survey in New Zealand

☐ - Fed's Framework Review: Jenny did a paper for it
↳ fig w/ LR-E (Hebden et al 2020)

- JMP 2 years ago: in cons. survey ppl are more aware of supply shocks: ⊕ corr btwn $E(u)$ & $E(\pi)$ vs. in firm/prof E: has much more demand shocks
- What forecasters report may differ from their beliefs b/c they may just report estimates from models which are stationary → lower bound.

☐ · Do you not get these mon. pol. results if cgain learning?

☐ Estimate a cgain regression using rolling reg. window
↳ and check whether cgain is corr. w/ fe and time-varying → gain isn't constant

Eq (10) A regression w/ interaction terms (fe & dummies)
→ rewritten form of the piecewise linear

A way to est (10) directly is to say s_t is a proxy : Principal components of data
OR : you do VAR-learning extension.

- CEMP isn't enough as evidence b/c for me the gain needs to be related to fe so I'd need to do that anyway.

☐ · bring ests back later in the talk
maybe when you intro the learning framework & you can talk how this is diff from other deviations ⟶ Anchoring mechanism slide: after.

☐ • Could also do my estimation w/ the constant gain case and you can show how much worse it fits the moments!

• My skillset is a good fit for the Board
→ frame as Policy Framework!

↳ Fabian Winkler (Board) LSE
↳ reach out!
he was also working on the Framework Review

↳ was Marcel there? one of the learning

↳ they do a lot of these learning models
→ need for ppl who get these models

Bos Fed Fellowship:

- 2 other students
- schedule a prezi
- we'll meet regularly: every other week, Fri 10.30.
- Tues 10.30-12pm. Seminar series.
  ↳ schedule 1-on-1 meetings w/ speakers typically on the afternoon after the seminar
  The first meetings will be Wed morning.

7 Sept 2020

To do

- [x] Label obs in PEA for opt & TR (quarters) → x axis
- [x] Label autocovariogram (lags) → x-axis
- [ ] optimal policy as a fct of k

~~Log-dev from st.st. on y-axis~~

→ [ ] estimation! figure out what's up w/ E

Read:

- Sargent, Conquest (at least Intro)
- Marvin Goodfriend & Bob King: The Incredible Volcker Defl.
- Bordalo et al (2x)
- Hebden et al

Sargent, Conquest, Chapter Intro

Two stories: "triumph of the natural rate hypothesis"
vs.
"vindication of econometric policy evaluation"

natural rate hypothesis := Phillips curve isn't exploitable in LR

triumph: "Solow & Samuelson told policymakers to exploit PC, which disappeared, Lucas' critique was internalized by policymaker."

vindication : "Policy was conducted according to Solow & Sammelson's suggestion, and this lead to policy learning the nat rate hypothesis.

He says that estimates in Chapter 9 confirm the vindication story.

↳ But his main concern is how policy learned to inflation-stabilize, i.e. how exactly the nat. rate hypothesis was internalized by the policymaker. It's a different question than mine.

Goodfriend & King, The Incredible Volcker Disinflation

Okun (1978): $\pi\downarrow$ can only come at the cost of high losses in output

Volcker disinflation wasn't as costly as predicted.

G&K : all the costs are attributable to the fact that agents perceived Volcker as <u>not</u> credible : they thought it likely that high inflation would return

↳ their model: imperfect credibility which implies very stubborn inflation expectations

Reading <u>Bordalo et al</u> <u>8 Sept 2020</u>

<u>Diagnostic Expectations & Credit Cycles</u> (finance)

<u>Over-reaction in macroeconomic expectations</u> (macro)

kind of together. But emphasis on the macro paper.

① Coibion & Goro (2015): "systematic departures from statistical optimality": forecast errors predictable

- H_0: FIRE. H_1: info rigidities. H_0 rejected

② Macro departures from FIRE
- RI (Sims 2003)
- bounded rat. (Gabaix 2014)
- info rigidities (Mankiw & Reis 2002, Woodford 2003)

Info rigidity: info costly → agents update sporadically → under-reaction to news which is what Coibion & Goro 2015 find

Finance: leading puzzle is over-reaction to news

Motivation of macro paper: can we account for both?

Part 1: document in data like CG.

reaction to news := fcst-revision (FREV)

$$fe_t = \beta_0 + \beta_1 FREV_t + \varepsilon_t$$

If $\beta_1 > 0$, underreaction

If $\beta_1 < 0$, overreaction

- consensus fcst (i.e. avg fcst across all SPF analysts)
  under-react
- indi fcsts over-react

How do get this pattern in one unified framework?

over-reaction ← representativeness heuristic

under-reaction in the agg: when each fcster over-reacts to private info, b/c that way they ignore public info, so they under-react to the average signal

↳ My framework can yield both depending on the size of the gain, which is state-dependent, just as CG found.

Sect 5.2. time-series test of diagnostic E

- most variables exhibit hump-shaped dynamics
- beliefs exaggerate not only on short-term
  ↳ in other words, oscillatory dynamics

Counter "noisy info + RE": ~~at~~ odds w/ how ppl behave in the lab (Cars Hommes)

Counter "diagnostic E": can match underreaction on agg & overreaction at indi level
but can't match the prevailing state-dependence of over- & underreaction

(which is in fact anchoring or unanchoring)

↳ all of these theories share w/ RE the implication that LR-E should be constant and in particular not vary systematically w/ fc.

Diagnostic E – the model

$$\omega_t = b\,\omega_{t-1} + e_t \qquad \varepsilon_t \sim N(0, \sigma^2)$$

fundamental, unobservable (macro paper)

macro paper: unobservability of ω_t (x_t) has 2 interpretations:

"noisy RE" ⟸ b/c both lead to optimal use of not fully available (noisy) info

- RI: too costly to observe x_t fully
- dispersed info: ppl have different bits of info

The agent must predict ω_{t+1} based on ω_t

The true distrib of ω_t is $h(\cdot)$. Instead, agents entertain the <u>relative</u> distrib $\frac{h(\cdot \mid G)}{h(\cdot \mid -G)}$

where "G" is the group and $-G$ is the comparison group ("context")

Representativeness of $\hat{\omega}_{t+1}$, a particular future state, is

$$\frac{h(\hat{\omega}_{t+1} \mid \omega_t = \hat{\omega}_t)}{h(\hat{\omega}_{t+1} \mid \omega_t = b\hat{\omega}_{t-1})}$$

i.e. a particular future state $\hat{\omega}_{t+1}$ is representative if it's relative likelihood given today's realization $\hat{\omega}_t$ has gone up relative to yesterday's, $\hat{\omega}_{t-1}$.

$$\Rightarrow h_t^{\theta}(\hat{\omega}_{t+1}) = h(\hat{\omega}_{t+1} \mid \omega_t = \hat{\omega}_t) \cdot \left[\frac{h(\hat{\omega}_{t+1} \mid \omega_t = \hat{\omega}_t)}{h(\hat{\omega}_{t+1} \mid \omega_t) = b\hat{\omega}_{t-1}}\right]^{\theta} \frac{1}{Z}$$

distorted distrib in agents' minds = true distrib $\cdot$ $\begin{pmatrix}\text{representativeness}\\ \text{distorting factor}\end{pmatrix}^{\theta}$ $\cdot$ normalizing constant

$\Rightarrow$ they take θ to be fixed, "however, recall can depend on the agent's deliberate effort, in which case θ may vary across situations" (finance p. 10 Mac)

↳ in my model, I allow agents to adopt their fcst framework to the sitn, θ is endogenous.

$$E_t^{\theta}(\hat{\omega}_{t+1}) = E_t(\omega_{t+1}) + \theta\left[E_t(\omega_{t+1}) - E_{t-1}(\omega_{t-1})\right] \quad (2)$$

diagnostic E = RE + θ (ΔRE-expectation)

↳ FREV!

What my model captures that diagnostic E don't:

1). Expectations deviate from RE by a constant factor (θ)

↳ in my case, this factor is time-varying, allowing the PS to respond to the volatility of the environment w/ its expectation-formation process

~~2.) LR-E~~. ← Shit. That's not true! See p. 13 Mar, eq(3).

2) Diagnostic E agree that there is a sense of lack of model knowledge b/c agents inflate the prob of representative states → but over time they should learn the true distribution

⇒ it presupposes a sense of static expectation formation in which agents cannot correct things about their E-formation that data indicates is untrue which conflicts w/ experiments of Hommes.

The average diagnostic forecast is rational. But how can it capture underreaction then on avg?

Since E(shocks) = 0, diag E = RE over time → the model does not display longer-term feedback loops, while mine does!

"noisy RE" → diagnostic E → anchoring

$fe \perp FREV$
b/c info friction
but optimal
use of info
(RI looks like this too)

[bounded rat. (Gabaix) too?]
$Corr(fe, FREV) \neq 0$
b/c agents overreact to what
they at the moment deem
representative

$Corr(fe, FREV)$
$\neq 0$ and
changes sign
based on
how E-formation
responds to the
environment

The macro paper is really an estimation horserace
meant to demonstrate that diagnostic E is good
positive model of E. It doesn't involve a model.

→ So Jenny must have meant the credit cycles paper
as having a violation of EH in the appendix.

→ Need to ass GARCH-process for fundamentals to get
θ to be time-dependent ⇒ at odds w/ lab experiments
showing that agents use simple rules.

Internet Appendix

When $-G$ is the past diagnostic expectation, then agents overreact on impact (at odds w/ data) and display reversals b/c the reference is moving too. But apparently the std model has this too?

It is inconsistent that agents don't learn the true degree of representativeness – a feature they try to address in App. B ("slow-moving - G")

↳ this is necessary to match initial underreaction to shocks. (Refer to Gennaioli, Shleifer & Vishny (2015))

Hebden et al 2020

↳ background to the Fed's Review of its Policy Framework

Main question: how robust are makeup strategies to alternative modeling assumptions, in particular about π-E(·)?

makeup strategies := policies that aim to offset past misses of π from its target

Main result: work best when the public understands, believes and reacts to the policymakers' commitment to offset. If expectations don't behave as desired, more aggressive policy accomodation is necessary (Goodfriend & King).

Motivation for makeup strategies: binding ZLB that forces the Fed to miss its current target.

The main makeup strategy: average inflation targeting (AIT)

A flatter PC has similar implications for makeup strategies as expectations that do not believe/understand the CB's policy commitment $\Rightarrow$ meeting the π-target then requires a larger output cost.

Empirical evidence on the anchoring of π-expectations.

Ways to measure:

(1) average agents' beliefs about π remaining within some range of the target

(2) low cross-sectional dispersion of forecasts around the target $\leftarrow$

(3) low subjective uncertainty in beliefs around the target
↳ measured from density forecasts

(4) small forecast revisions and little response of revisions to news

Ryan meeting 9 Sep 2020

- Susanto: a google sheet w/ open positions
  next deadline: Sep 30. → Our convo? - Next week and he'll ask to see revealed preferences
- GLMM
- Dawid's feedback (maybe)

- Check whether loss improves from Fig 3 & 4.
- Edges don't matter much b/c fe of 3 or 4 don't occur so much.
  ↳ Fig 3 is already good!

Fig 7 Space btwn blue & red too big between Exp - moments → I didn't annualize $\bar{\pi}$ in the true synthetic data!

→ focus on $Var(E(\bar{\pi}))$
line up seeds, $N=1$ and look at history to see where they diverge.

Draft to Ryan: Sept 21.

Work after

→ now w/ annualized $\bar{\pi}$ they are identical!

Peter meeting

10 Sept 2020

(1) • Susanto : sheet of open positions
(2) Ryan : tab w/ talks
(3) ↳ Do you wanna discuss positions? → We'll talk.
(4) → Do you censor at the top ; meaning are there too good positions you think I shouldn't even apply for? → No censoring :)
(5) • Ryan : draft 21 Sept.

Federico Mandelman Atlanta Fed

U Mississippi → Peter has friends there
(Oxford)

He went on JM 1991 (1st Iraq War) (he was 25 ...)
Offers: U Texas Austin
Texas A&M
Richmond & K.C. Fed

Biweekly 1.30pm Thurs (2 weeks from now)