Materials 38 - Bias (not only?) in the neighborhood of zero forecast errors

Laura Gáti

July 28, 2020

Overview

1 Ways to try to get identification

3 potential causes to lack of identification in the zero neighborhood

1. The distribution of estimates is skewed $\rightarrow$ take $median(\hat{\alpha})$ instead of the mean.

   $\Rightarrow$ also unidentified

2. The gain doesn't matter if the forecast error is 0, or very close to it $\rightarrow$ introduce a distinction between the forecast error that's used to choose the gain and the one used to update the coefficients of the learning rule.

   $\Rightarrow$ tried using a different forecast error (time, output gap), also unidentified

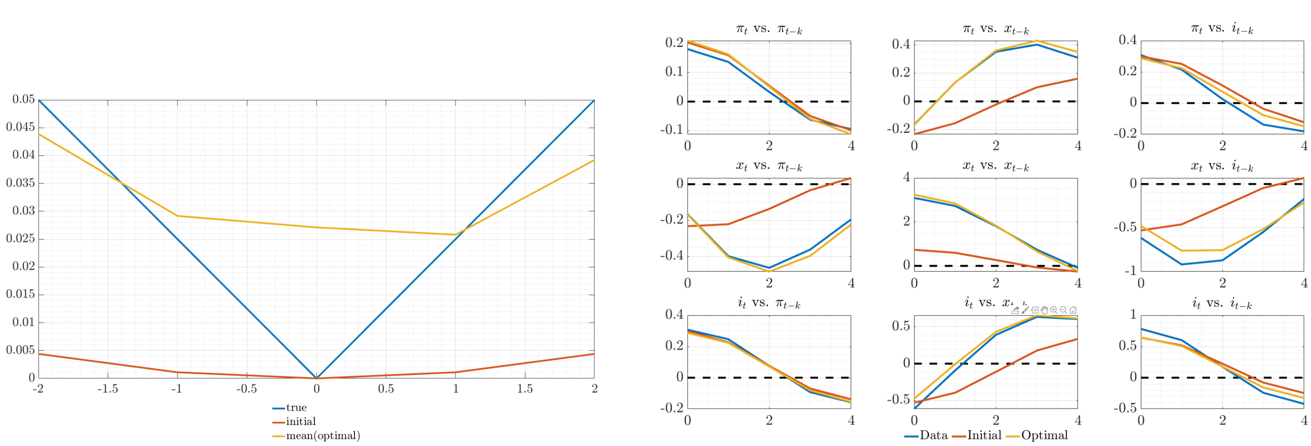
3. Introduce expectation series (SPF)

4. Taking mean moments across N histories instead of performing the estimation N times.

5. The truth is based on a simulation that doesn't favor the zero neighborhood $\rightarrow$ do 100 simulations from the "true" parameters and take the mean moments of those.

Reference for comparison: Fig 1. of Materials 37

Figure 1: Reference figure: Mean estimates for $N = 100$, imposing convexity with weight 100K, truth with $nfe = 5, fe \in (-2, 2)$



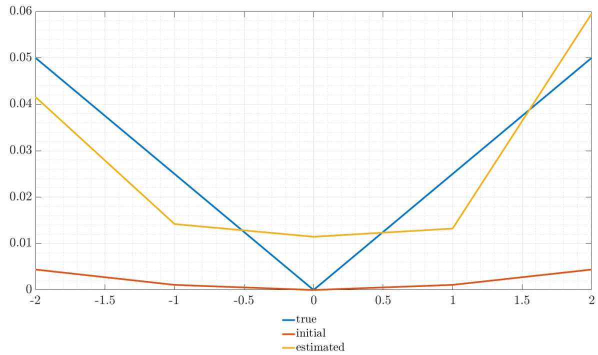
(a) $\alpha^{true}, \alpha_0, mean(\hat{\alpha})$

(b) Autocovariogram

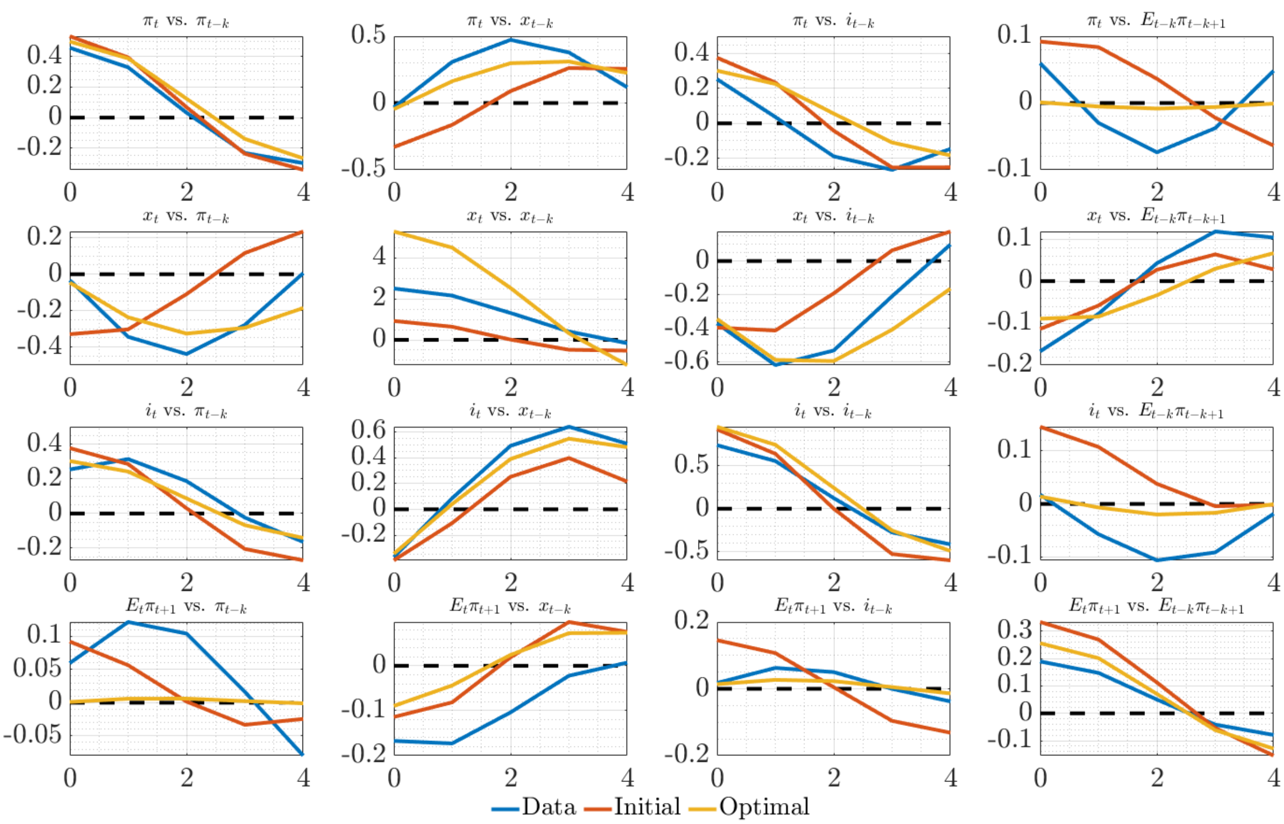
1.1 Point 3: add expectations series

I've added measurement error to π, x, i and the expectation to avoid stochastic singularity from having 4 observables and only 3 shocks.

Figure 2: Estimates for $N = 100$, incl. 1-step ahead forecasts of inflation, imposing convexity with weight 100K, truth with $nfe = 5, fe \in (-2, 2)$



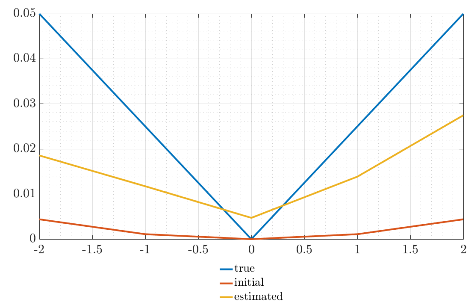
(a) $\alpha^{true}, \alpha_0, mean(\hat{\alpha})$



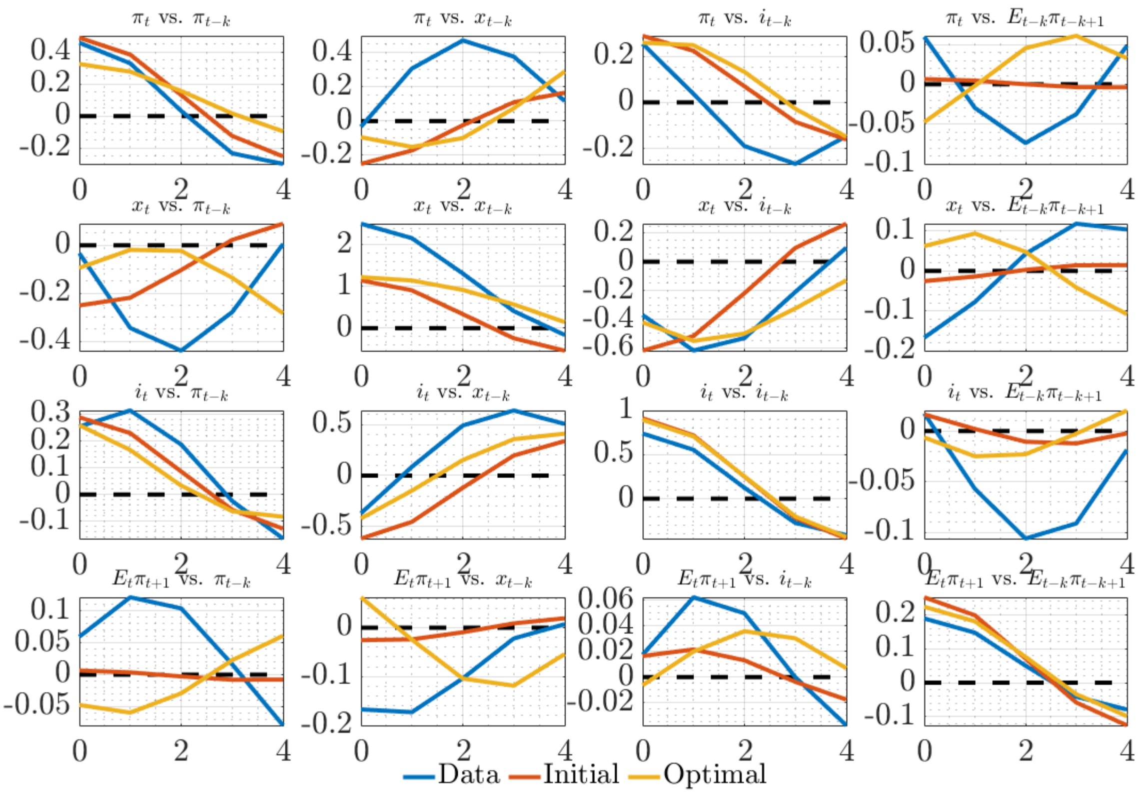
(b) Autocovariogram

1.2 Point 4: Mean moments instead of N estimations

Figure 3: Estimates for $N = 100$, incl. 1-step ahead forecasts of inflation, imposing convexity with weight 100K, truth with $nfe = 5, fe \in (-2, 2)$, single estimation of mean moments from N simulations



(a) $\alpha^{true}, \alpha_0, mean(\hat{\alpha})$

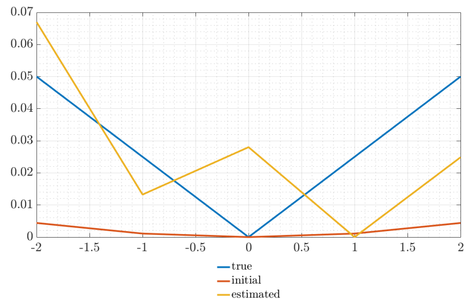


(b) Autocovariogram

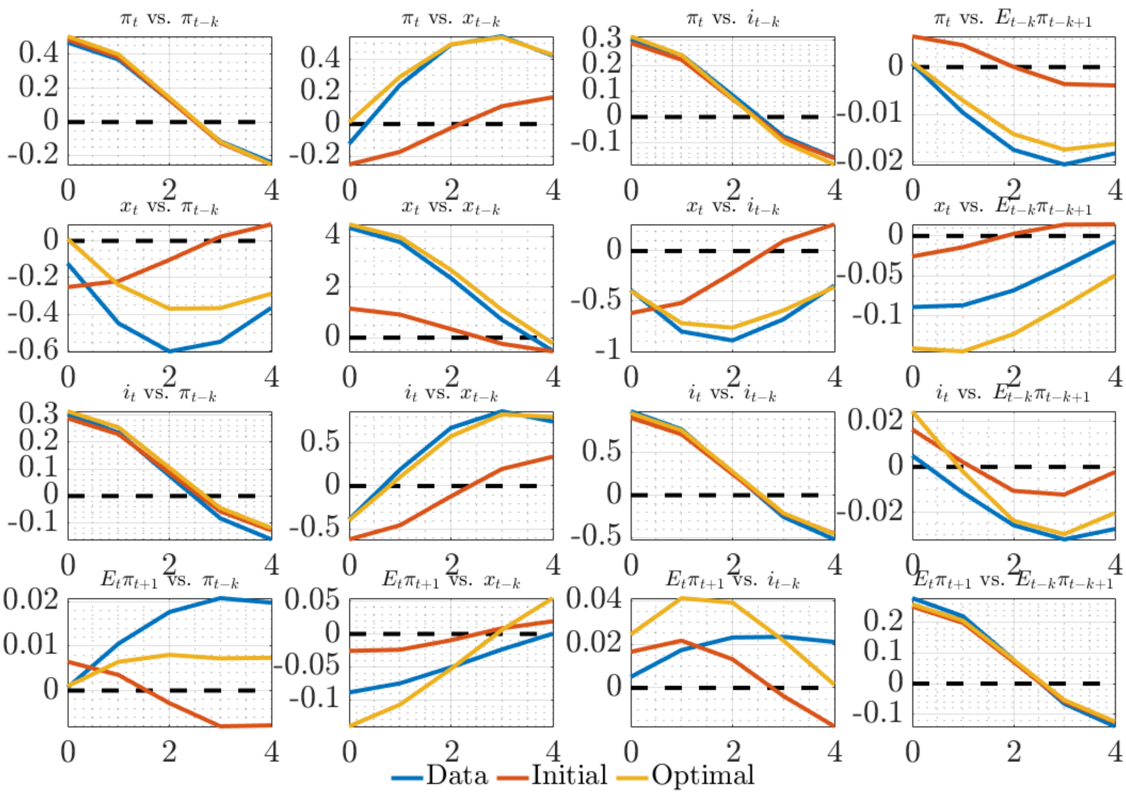
This does make a difference and I think it improves on the moments vis-a-vis the N estimations case. However, it converges in the wrong direction with $N = 1000$. And it really depends on shocks! A seed of `rng(2)` instead of `rng(1)` makes a huge difference at $N = 100$, b/c $N = 100$ doesn't seem sufficient to wash out the shocks. $N = 1000$ seems sufficient though. The "N-estimations" strategy however is robust to changing the seed.

1.3 Point 5: 100 simulations from truth

Figure 4: Estimates for $N = 100$, incl. 1-step ahead forecasts of inflation, imposing convexity with weight 100K, truth with $nfe = 5, fe \in (-2, 2)$, single estimation of mean moments from N simulations, 100 truths



(a) $\alpha^{true}, \alpha_0, mean(\hat{\alpha})$



(b) Autocovariogram

Didn't converge!

1.4 Does the estimation strategy work?

Try in a simplified setting with

$$s_t = 0.7 s_{t-1} + \epsilon_t \tag{1}$$

$$y_t = \alpha b(s_t) \qquad \text{piecewise linear approx of state} \tag{2}$$

Estimate the same α^{true} as before, filtering the same way, calculating 5 moments (autocovariances of y) the exact same way.

- The "estimate mean moments" strategy is not robust (gets different, incorrect things depending on seed or initialization, even for 2 knots), just like for my application.
- The "estimate N times" procedure is biased but robust just like in my application: biased everywhere but most in the middle (even for 2 knots)).

⇒ Is there something about the way I compute the moments that renders them uninformative?

1.4.1 How I compute the moments

I first fit a VAR to the BK-filtered data (I use p I estimated in the data, usually 4). Then I rely on Hamilton, p. 266 on (p. 280 Mac), Notes 12, p. 25, to estimate the autocovariances of a VAR as:

1. Rewrite the estimated n-variable VAR(p) as a VAR(1) as:

$$\xi_t = F\xi_{t-1} + v_t \tag{3}$$

   and define the VC matrix of the ξ as $\Sigma \equiv \mathbb{E}(\xi_t \xi_t')$. The idea is to estimate Σ from the VAR(1) representation and then back out the submatrices we need.

2. Take the square of (3) and take expectations:

$$\Sigma = F\,\mathbb{E}(\xi_{t-1}\xi_{t-1}')F' + \underbrace{\mathbb{E}(v_t v_t')}_{\equiv Q} \tag{4}$$

3. The solution to this is (Hamilton's equation [10.2.18]):

$$vec(\Sigma) = [I_{(np)^2} - F \otimes F]^{-1} vec(Q) \tag{5}$$

4. The jth autocovariance of the process y_t is given by the first n rows and n columns of Σ_j:

$$\Sigma_j = F\Sigma_{j-1} \quad = F^j \Sigma \qquad \text{(Hamilton's eq. [10.2.20]) and [10.2.21])} \tag{6}$$

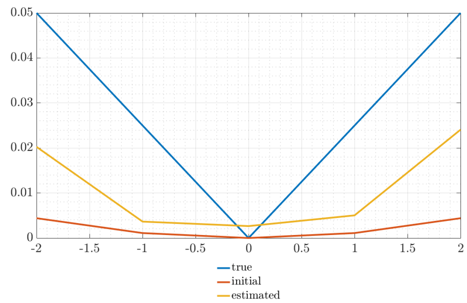
I've also tried to HP-filter instead of BK-filtering, for a truth with only 2 knots, both mean moments and N estimations strategies, results unchanged.

One thing I am noticing is that the estimated variance of the moments is very tiny, but only for the synthetic data, not for the real data. This means that the weighting matrix of the quadratic form is on the order of between 10^4 and 10^{10} (lower if higher number of knots). Can this ruin the informativeness of the moments? (For my model, using an identity matrix substantially improves the matching of the moments.) And why is this only happening for synthetic data?

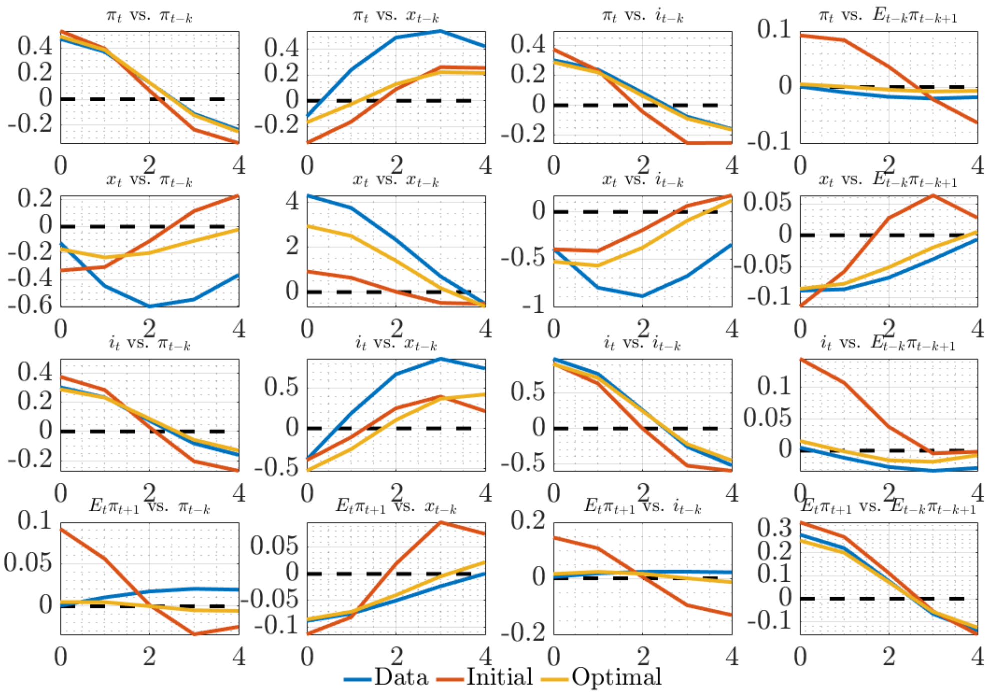
Replacing the piecewise linear approx with a linear rule with a slope coefficient works, but an intercept and a slope coefficient doesn't work to estimate the two parameters.

1.5 Rescale the GMM weighting matrix

Figure 5: Estimates for $N = 100$, incl. 1-step ahead forecasts of inflation, rescale W, imposing convexity with weight 100K, truth with $nfe = 5, fe \in (-2, 2)$



(a) $\alpha^{true}, \alpha_0, mean(\hat{\alpha})$

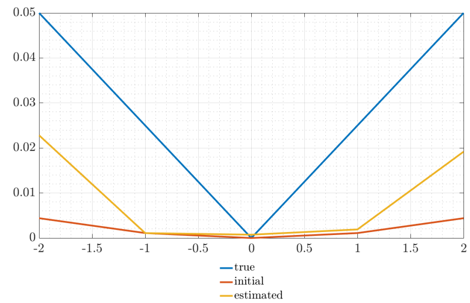


(b) Autocovariogram

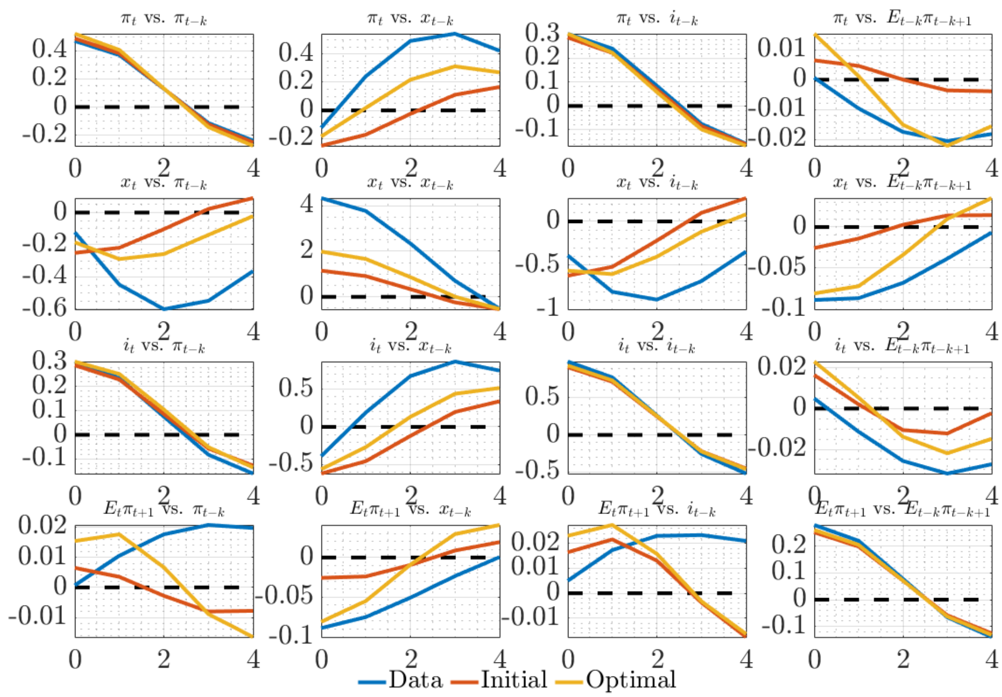
I take the smallest diagonal element of W, the variance matrix of moments. I then rescale W by the order of magnitude of the smallest element. That doesn't seem to do really well, although it does hit some moments better, others worse. It definitely stays closer to the initial values.

Also here we see it sticking to the initialization clearly.

Figure 6: Estimates for $N = 100$, incl. 1-step ahead forecasts of inflation, rescale W, estimate mean moments once, imposing convexity with weight 100K, truth with $nfe = 5, fe \in (-2, 2)$



(a) $\alpha^{true}, \alpha_0, mean(\hat{\alpha})$



(b) Autocovariogram

2 For me: Simulated "true" data

3 potential causes to lack of identification in the zero neighborhood

1. The distribution of estimates is skewed → take $median(\hat{\alpha})$ instead of the mean.

2. The truth is based on a simulation that doesn't favor the zero neighborhood → do 100 simulations from the "true" parameters and take the mean moments of those.

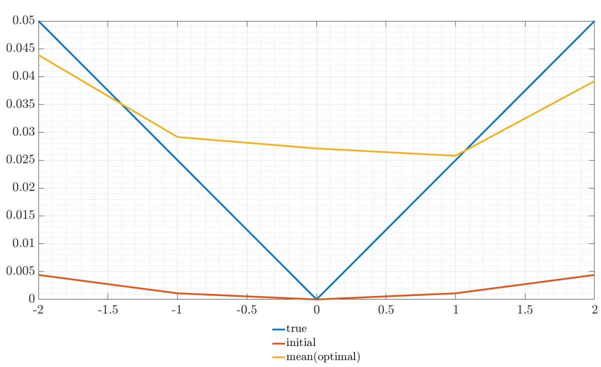
3. The gain doesn't matter if the forecast error is 0, or very close to it → introduce a distinction between the forecast error that's used to choose the gain and the one used to update the coefficients of the learning rule.

+1 Taking mean moments across N histories is more natural than performing the estimation N times.

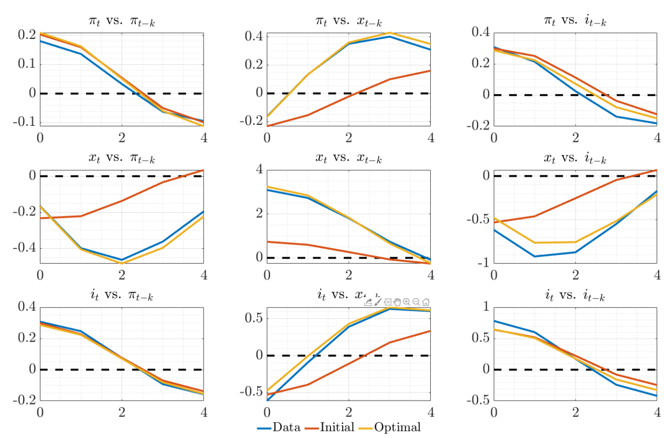
+2 Introduce expectation series (SPF)

Reference for comparison: Fig 1. of Materials 37

Figure 7: Reference figure: Mean estimates for $N = 100$, imposing convexity with weight 100K, truth with $nfe = 5, fe \in (-2, 2)$



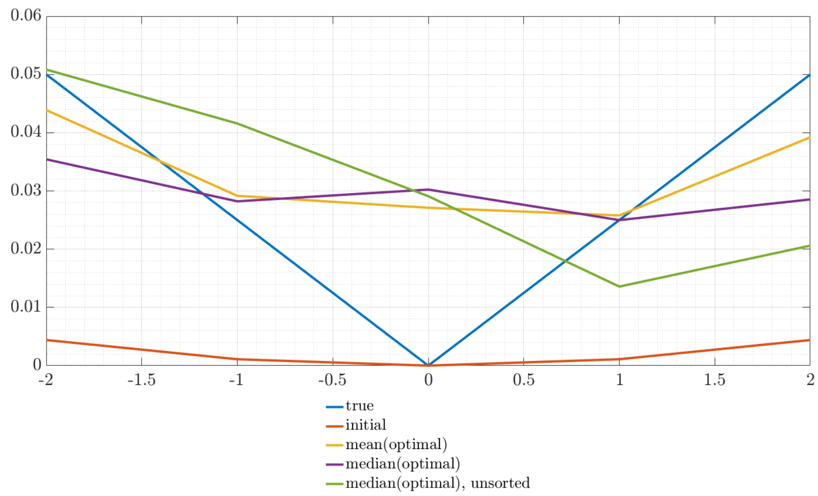
(a) $\alpha^{true}, \alpha_0, mean(\hat{\alpha})$



(b) Autocovariogram

Point #1: skewness → take median instead of mean

Figure 8: Mean estimates for $N = 100$, imposing convexity with weight 10K, truth with $nfe = 5, fe \in (-2, 2)$

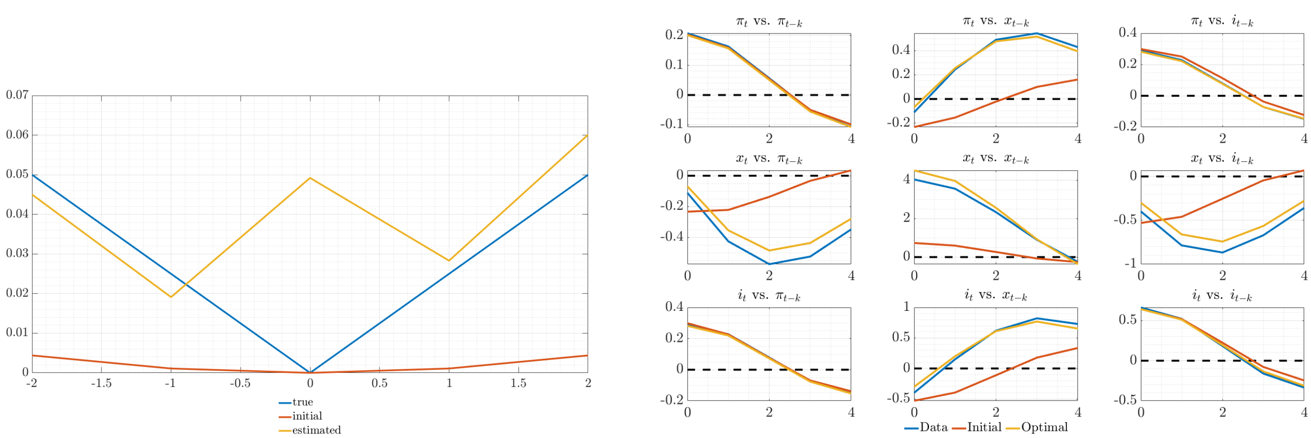


(a) $\alpha^{true}, \alpha_0, mean(\hat{\alpha}), median(\hat{\alpha})$, unsorted median

I understand what's happening! Half the estimates are L's, the other half are "inverted L's", which is why taking a mean or a classical, sorted median has the tendency to produce these nonmonotonic zigzags.

Point #2: do 100 truths

Figure 9: Estimates for $N = 100$, truth is a mean of 100 simulations, imposing convexity with weight 100K, truth with $nfe = 5, fe \in (-2, 2)$



(a) $\alpha^{true}, \alpha_0, mean(\hat{\alpha})$

(b) Autocovariogram

That didn't help, did it now?

Point #3: change timing of forecast errors

$$k_t^{-1} = \mathbf{g}(fe_{t|t-1}) \tag{7}$$

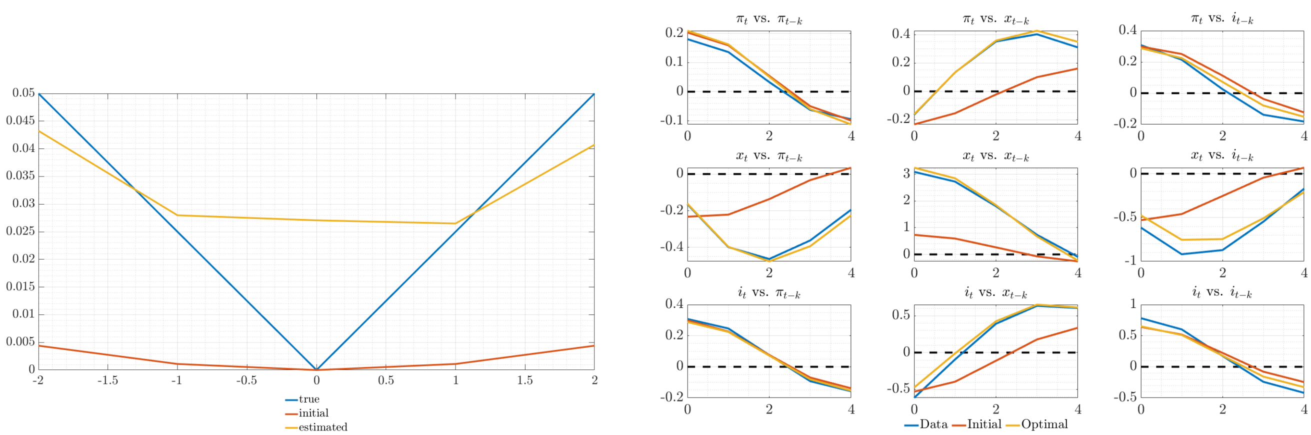
$$\bar{\pi}_t = \bar{\pi}_{t-1} + k_t^{-1} fe_{t|t-1} \tag{8}$$

The issue seems to be: if $fe_{t|t-1} \approx 0$, then the gain is irrelevant for learning because $fe_{t|t-1}$ figures into both equations. So the idea is to decouple the two equations by changing the timing of one of the forecast errors. Note:

$$\begin{aligned} fe_{t|t-1} &= \pi_t - (\bar{\pi}_{t-1} + bs_{t-1}) && (9) \\ &= \pi_t - \bar{\pi}_{t-1} \qquad \text{since shocks iid and } b \text{ is the RE transition matrix} && (10) \end{aligned}$$

So what I can try is to use an older forecast error in equation (2). Try $fe_{t|t-1} \equiv \pi_t - \bar{\pi}_{t-2}$.

Figure 10: Estimates for $N = 100$, changing the forecast error timing in the updating equation, imposing convexity with weight 100K, truth with $nfe = 5, fe \in (-2, 2)$



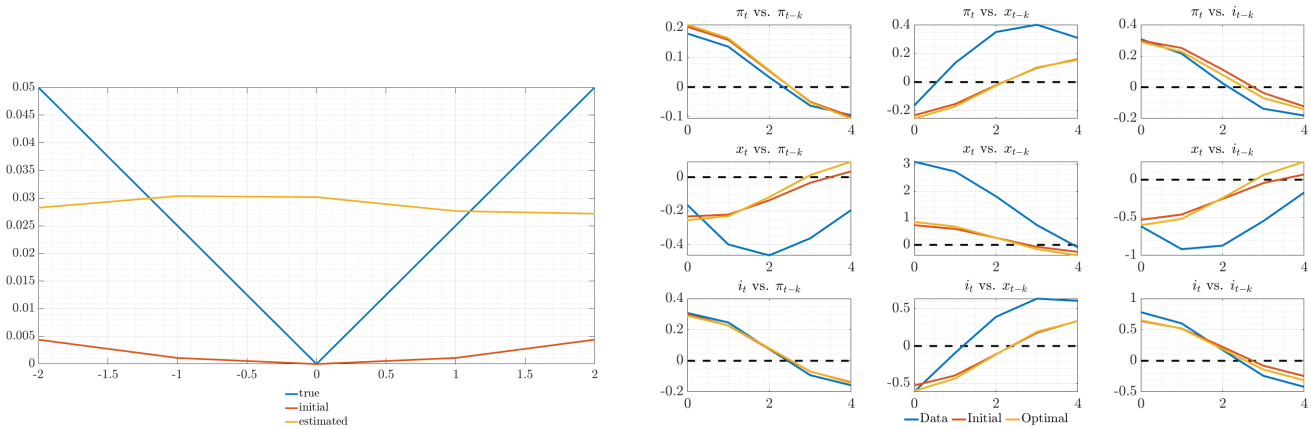
(a) $\alpha^{true}, \alpha_0, mean(\hat{\alpha})$

(b) Autocovariogram

A little more symmetric, but no dramatic improvement.

Point #+1: do N simulations instead of N estimations

Figure 11: Estimates for $N = 100$, targeting mean moments in a single estimation instead of N estimations of individual moments, imposing convexity with weight 100K, truth with $nfe = 5, fe \in (-2, 2)$



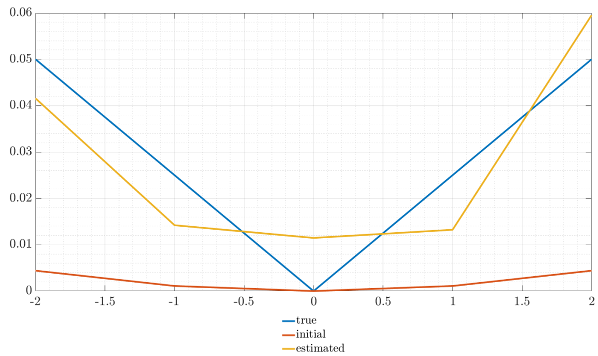
(a) $\alpha^{true}, \alpha_0, mean(\hat{\alpha})$

(b) Autocovariogram

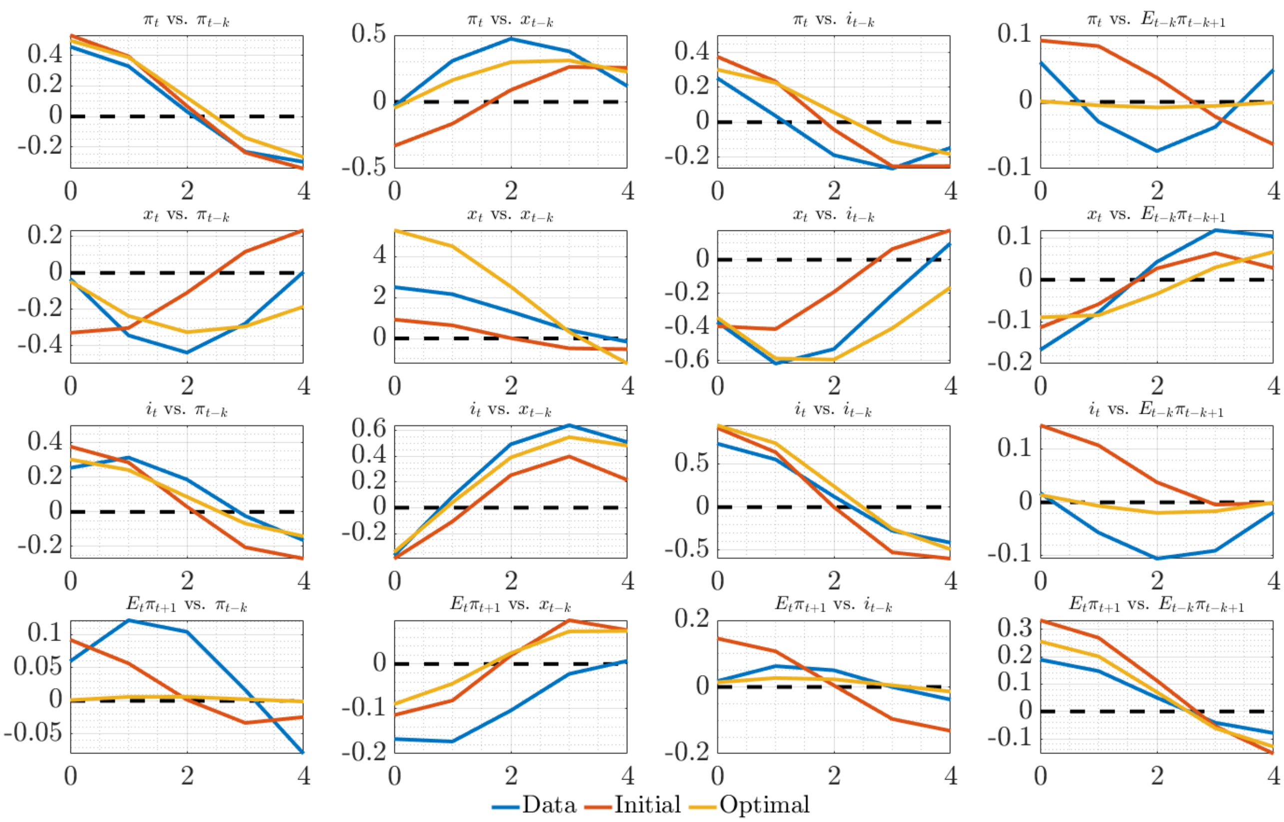
The difference is striking!

Point #+2: introduce expectations series

Figure 12: Estimates for $N = 100$, incl. 1-step ahead forecasts of inflation, imposing convexity with weight 100K, truth with $nfe = 5, fe \in (-2, 2)$



(a) $\alpha^{true}, \alpha_0, mean(\hat{\alpha})$



(b) Autocovariogram

I've added measurement error to π, x, i and the expectation to avoid stochastic singularity from having 4 observables and only 3 shocks.

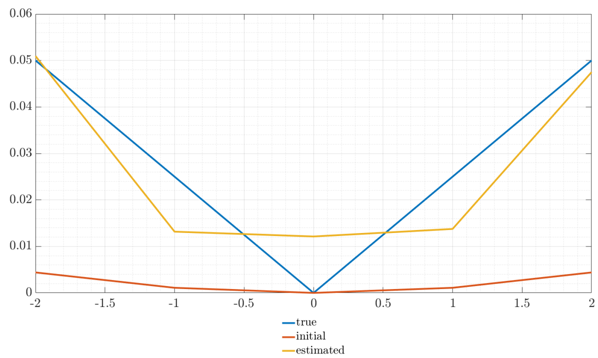
This clearly has added useful info. Otherwise, behavior is like before:

- Without the convexity restriction, I still get nonconvex estimate.

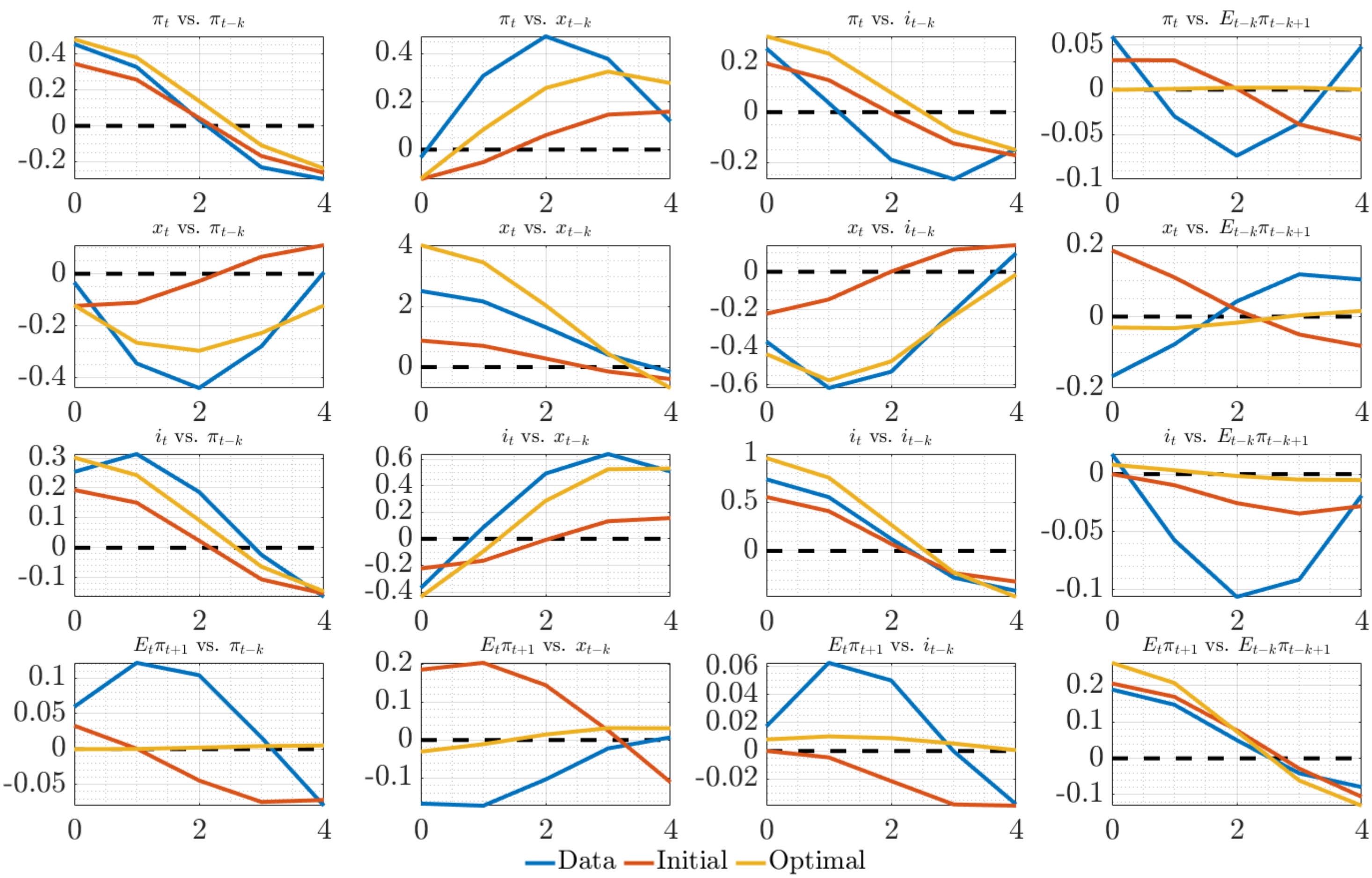
- With the 0 at 0 restriction, I can match the 0 region, otherwise I can't.

- Both restrictions lead to basically identical moments.

Figure 13: Estimates for $N = 1000$, incl. 1-step ahead forecasts of inflation, imposing convexity with weight 100K, truth with $nfe = 5, fe \in (-2, 2)$



(a) $\alpha^{true}, \alpha_0, mean(\hat{\alpha})$



(b) Autocovariogram

What's missing

To me it seems that the reason we're still underidentified is that the $(-1, 1)$-forecast error region (possibly an even bigger, $(-1.5, 1.5)$-region) doesn't produce variation in $\bar{\pi}$ (see Eq. 8). So let's try to replace the forecast error in the generation of the gain (Eq. 7) by say the forecast error of the output gap. (Needs the constant-only PLM.) → Doesn't work at all, I guess b/c it doesn't correspond to the DGP.

→ Really wonder if the anchoring function specified in terms of changes, not levels of the gain would help! If instead of equation 7 in system 7-8, we'd have

$$k_t^{-1} = \mathbf{g}(k_{t-1}^{-1}, fe_{t|t-1}) \tag{11}$$

No, even that wouldn't work b/c the estimation routine wouldn't be able to discriminate between two points on the line $(k_{(i)}^{-1}, fe \in (-1, 1))$, for $\forall i$ in the k-space.

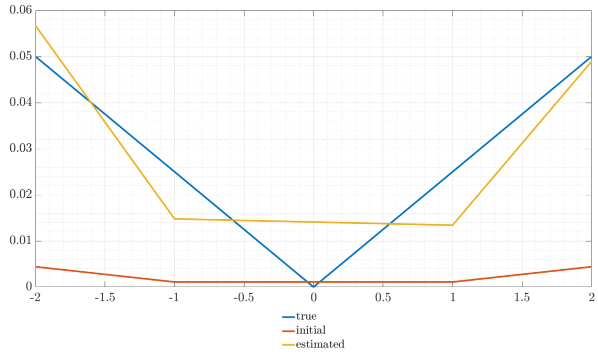
→ I'm increasingly thinking that that region cannot be identified at all because it simply doesn't matter for the evolution of long-run expectations.

2 options:

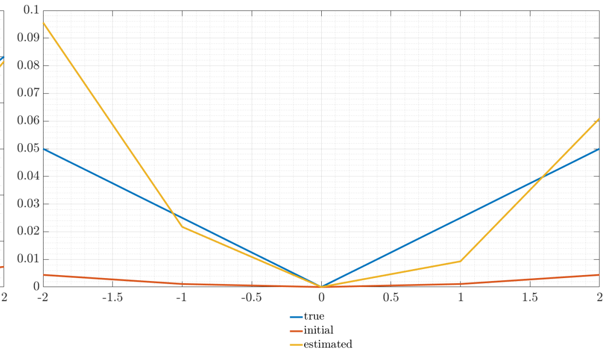
1. Select gridpoints strictly outside the $(-1, 1)$-region, and use the convexity restriction to interpolate inside the region.
2. Impose the 0 at 0 restriction at the point at 0, and select the rest of the gridpoints outside the $(-1, 1)$-region.

If my conjecture is correct, both of these approaches should be identified.

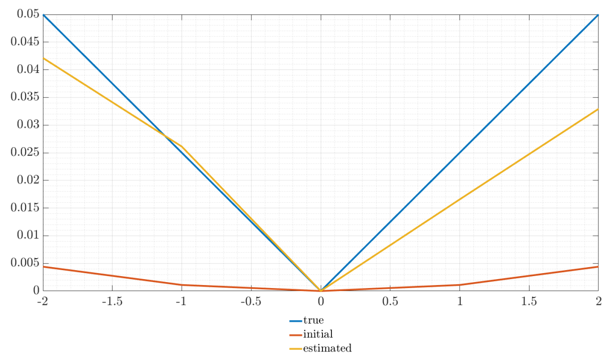
Figure 14: Estimates for $N = 100$, incl. 1-step ahead forecasts of inflation, imposing convexity with weight 100K, truth with $nfe = 5, fe \in (-2, 2)$



(a) $\alpha^{true}, \alpha_0, mean(\hat{\alpha})$, 4 knots: (-2,-1,1,2)



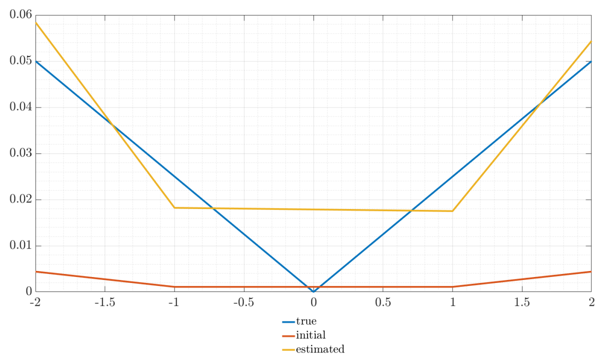
(b) $\alpha^{true}, \alpha_0, mean(\hat{\alpha})$, 5 knots: (-2,-1, 0, 1,2), 0 at 0 restriction imposed with weight 1K



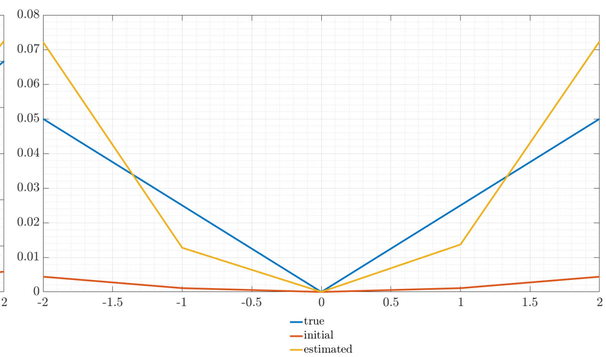
(c) $\alpha^{true}, \alpha_0, mean(\hat{\alpha})$, 5 knots: (-2,-1, 0, 1,2), 0 at 0 restriction imposed with weight 1K, convexity restriction removed

Unfortunately, it seems to me like even outside the trouble region, we're not identified. I say that because the 0 at 0 assumption on its own should have no bearing on the coefficients out in the tails of the forecast error space. But also there, even with a high N, I'm not nailing the truth.

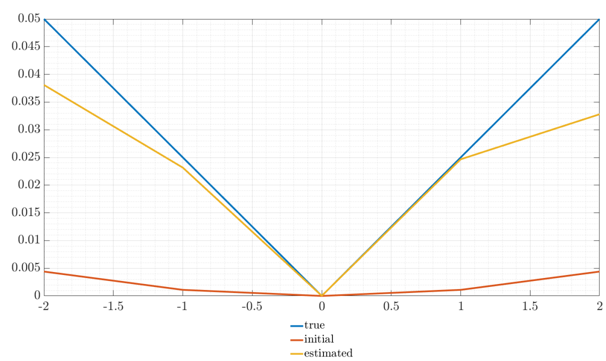
Figure 15: Estimates for $N = 1000$, incl. 1-step ahead forecasts of inflation, imposing convexity with weight 100K, truth with $nfe = 5, fe \in (-2, 2)$



(a) $\alpha^{true}, \alpha_0, mean(\hat{\alpha})$, 4 knots: (-2,-1,1,2)

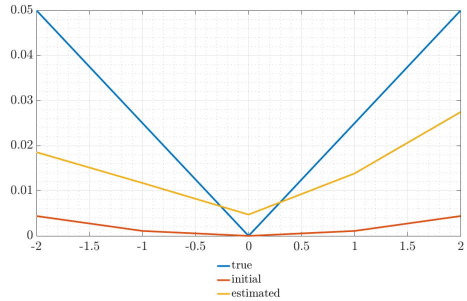


(b) $\alpha^{true}, \alpha_0, mean(\hat{\alpha})$, 5 knots: (-2,-1, 0, 1,2), 0 at 0 restriction imposed with weight 1K

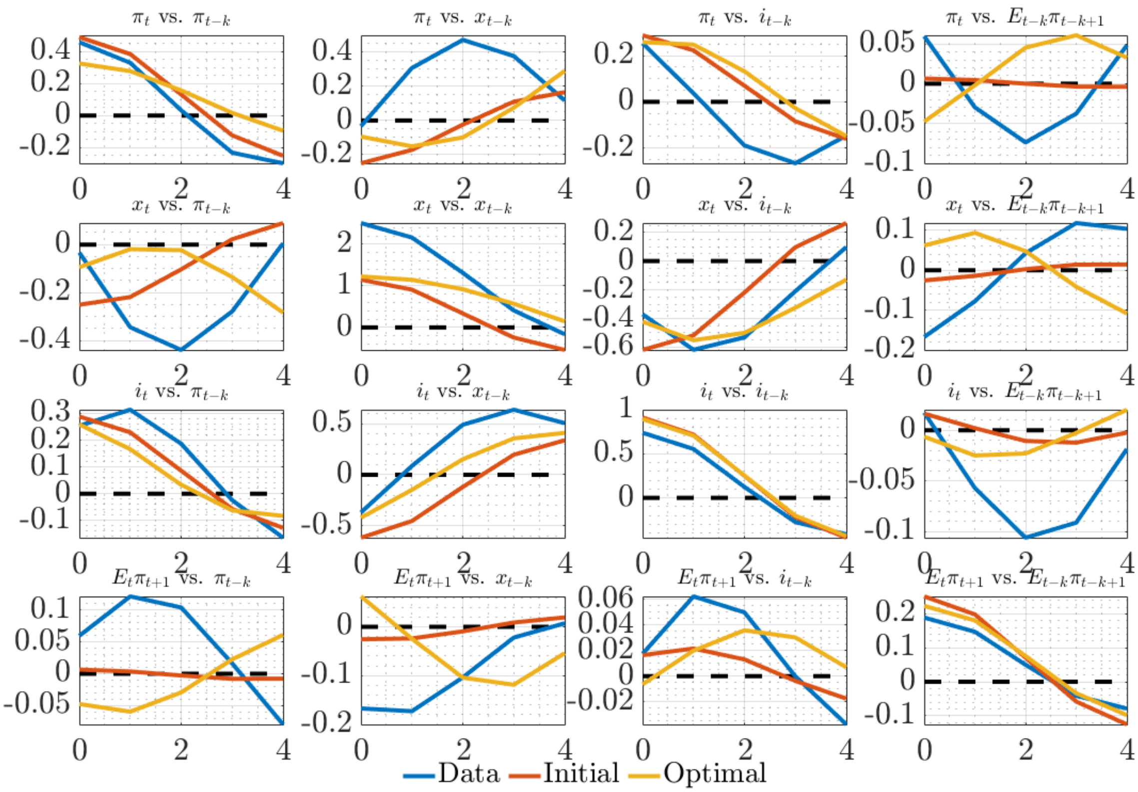


(c) $\alpha^{true}, \alpha_0, mean(\hat{\alpha})$, 5 knots: (-2,-1, 0, 1,2), 0 at 0 restriction imposed with weight 1K, convexity restriction removed

Figure 16: Estimates for $N = 100$, incl. 1-step ahead forecasts of inflation, imposing convexity with weight 100K, truth with $nfe = 5, fe \in (-2, 2)$, single estimation of mean moments from N simulations



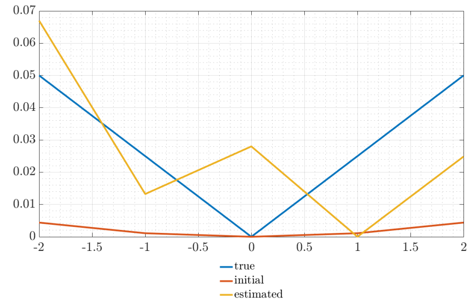
(a) $\alpha^{true}, \alpha_0, mean(\hat{\alpha})$



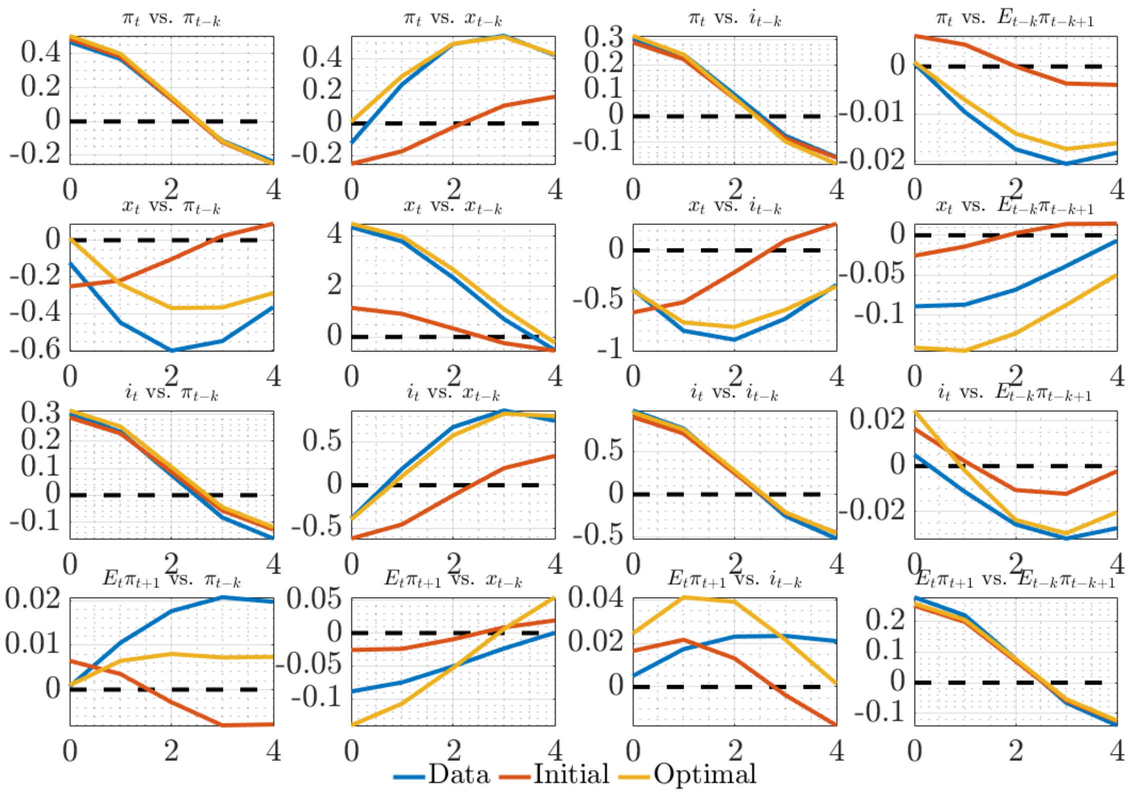
(b) Autocovariogram

This does make a difference and I think it improves on the moments vis-a-vis the N estimations case. However, it converges in the wrong direction with $N = 1000$. And it really depends on shocks! A seed of `rng(2)` instead of `rng(1)` makes a huge difference at $N = 100$, b/c $N = 100$ doesn't seem sufficient to wash out the shocks. $N = 1000$ seems sufficient though. The "N-estimations" strategy however is robust to changing the seed.

Figure 17: Estimates for $N = 100$, incl. 1-step ahead forecasts of inflation, imposing convexity with weight 100K, truth with $nfe = 5, fe \in (-2, 2)$, single estimation of mean moments from N simulations, 100 truths



(a) $\alpha^{true}, \alpha_0, mean(\hat{\alpha})$

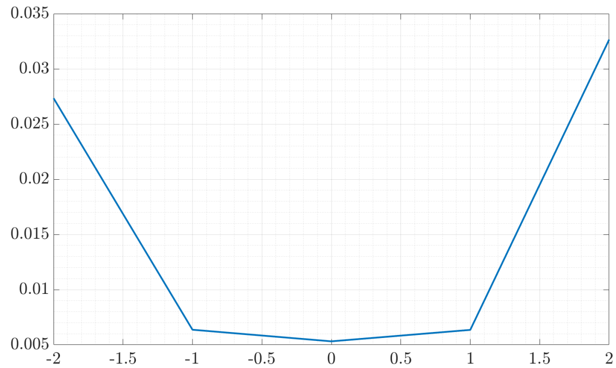


(b) Autocovariogram

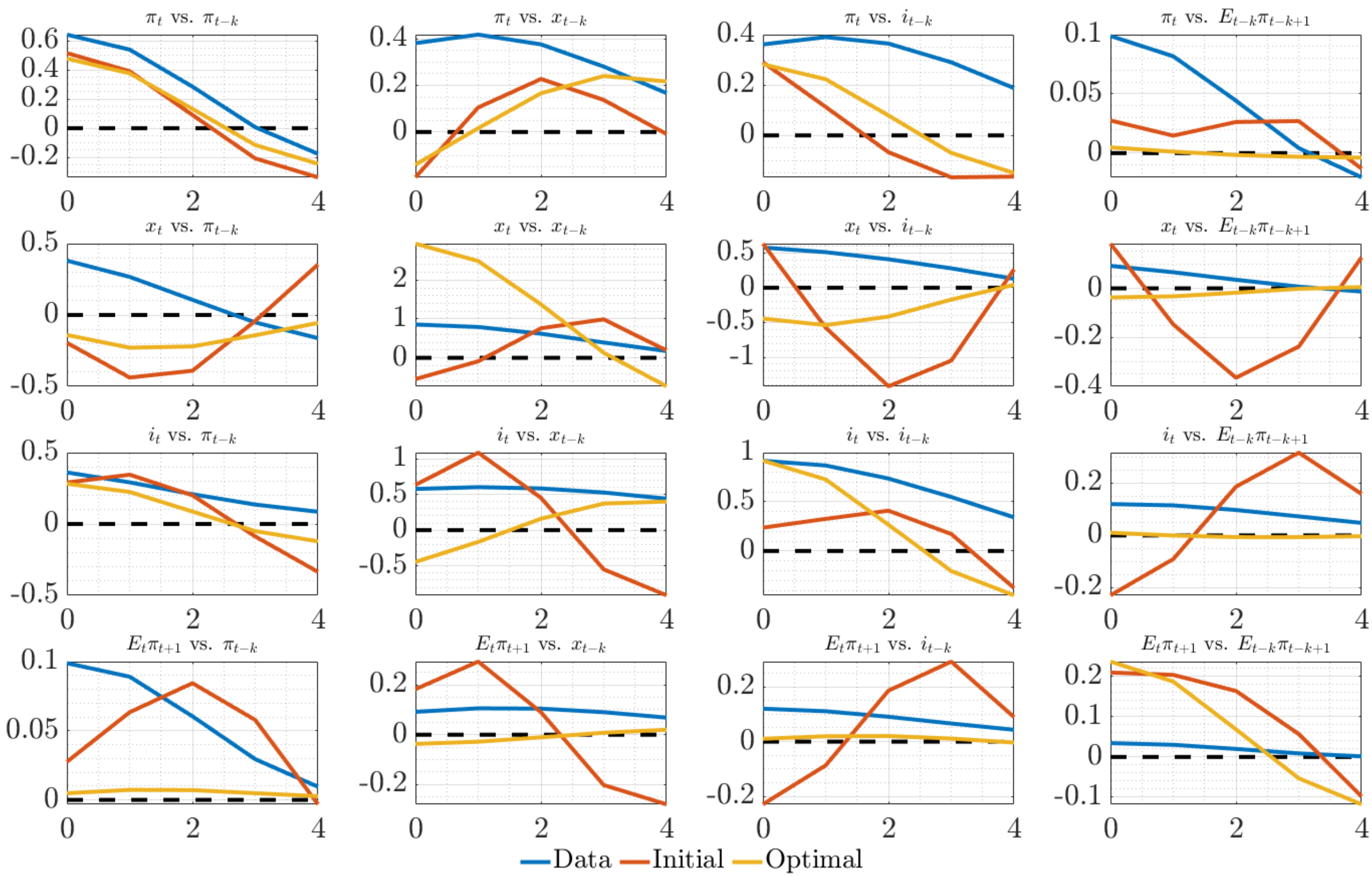
Didn't converge!

3 Real data with the SPF

Figure 18: Estimates for $N = 1000$, incl. SPF 1-step ahead forecasts of inflation, imposing convexity with weight 100K, $nfe = 5, fe \in (-2, 2)$

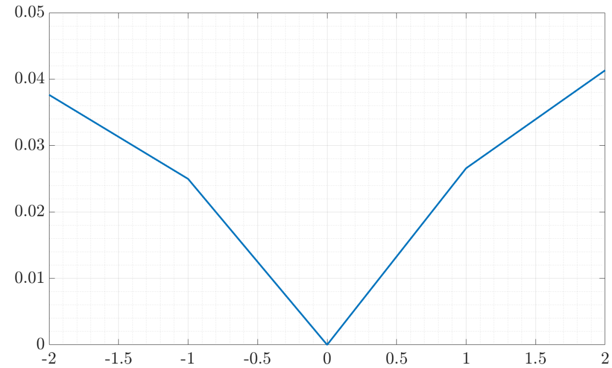


(a) $\hat{\alpha}$

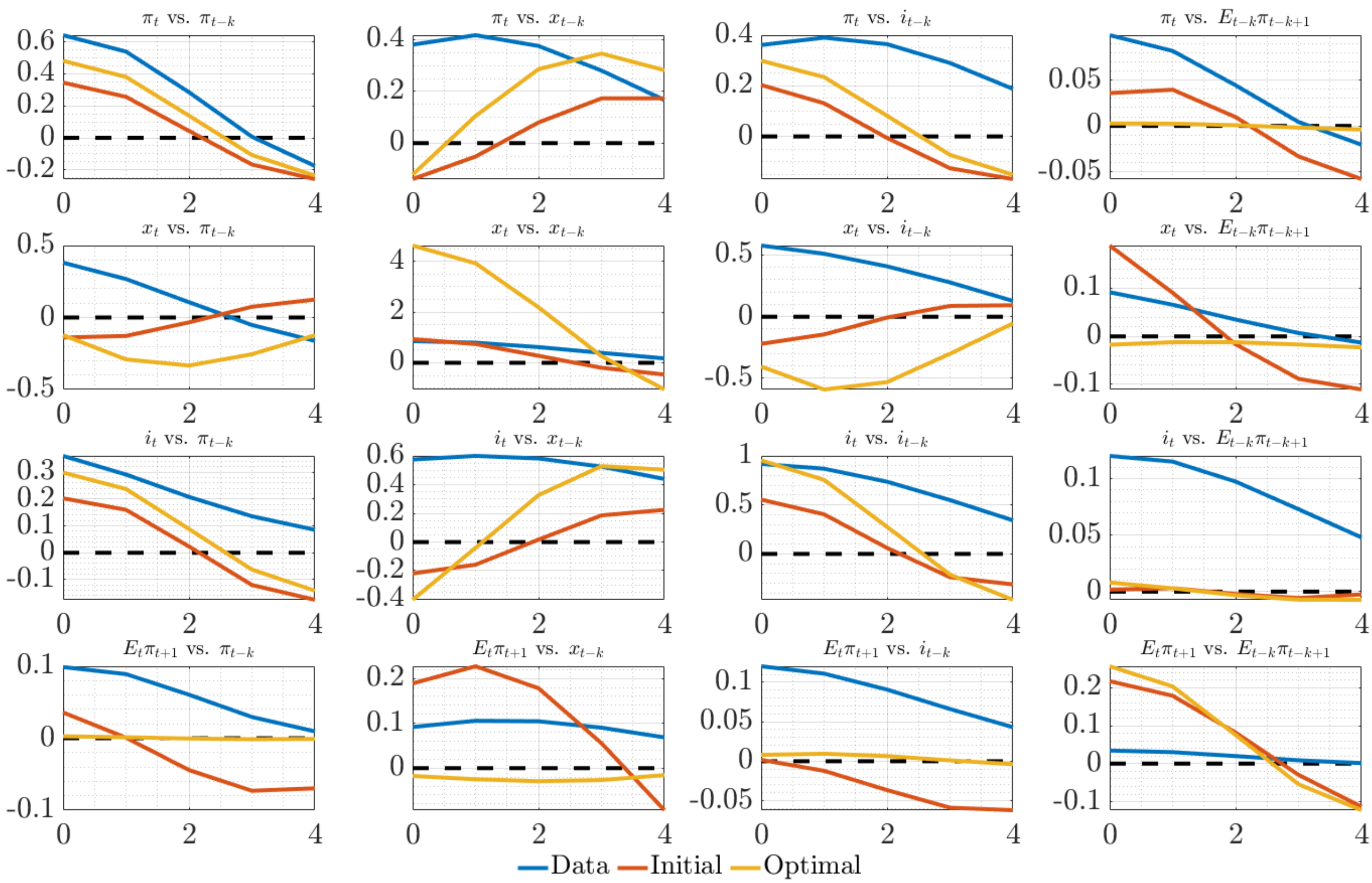


(b) Autocovariogram

Figure 19: Estimates for $N = 1000$, $nfe = 5, fe \in (-2, 2)$, incl. SPF 1-step ahead forecasts of inflation, removing convexity restriction, imposing 0 at 0 with weight 1K

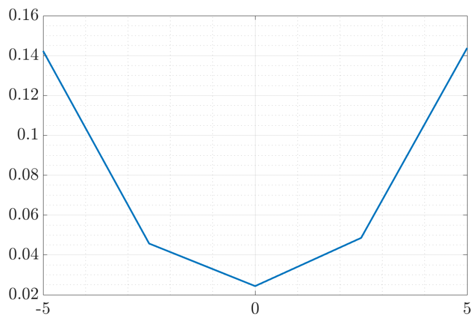


(a) $\hat{\alpha}$

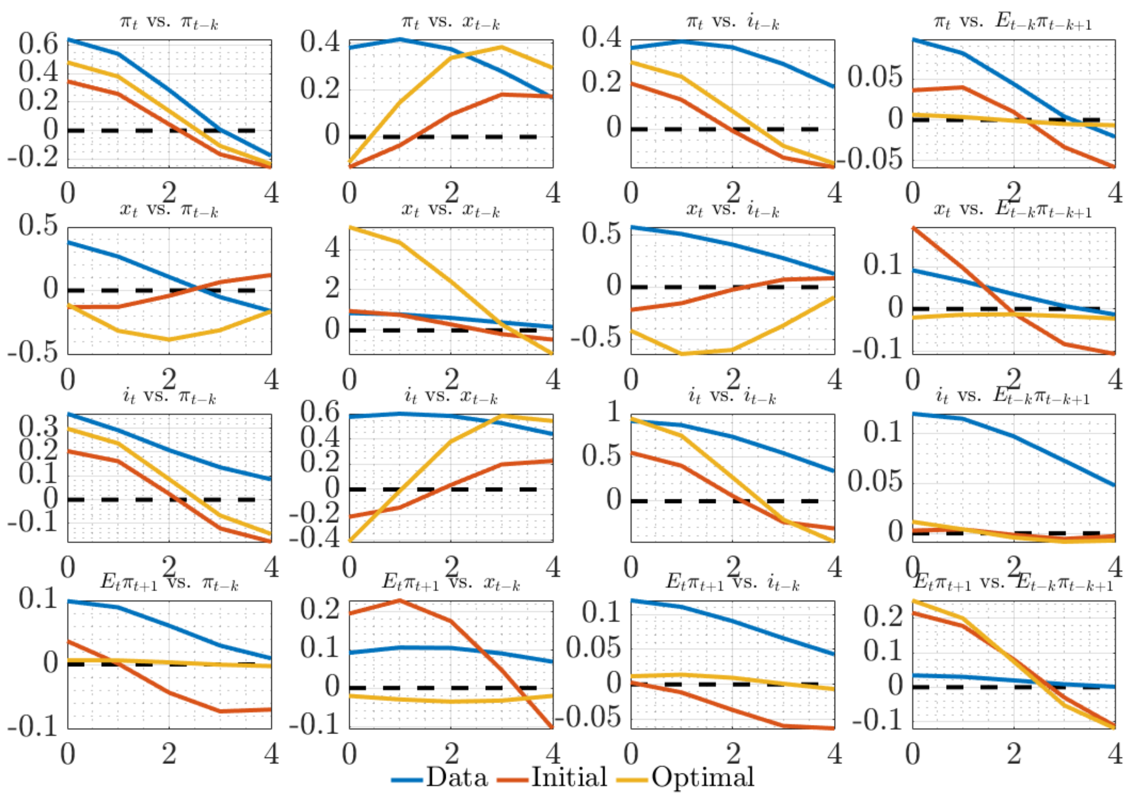


(b) Autocovariogram

Figure 20: Estimates for $N = 1000$, incl. SPF 1-step ahead forecasts of inflation, imposing convexity with weight 100K, 5 knots, $fe \in (-2, 2)$



(a) $\hat{\alpha}$



(b) Autocovariogram

A Model summary

$$x_t = -\sigma i_t + \hat{\mathbb{E}}_t \sum_{T=t}^{\infty} \beta^{T-t}\big((1-\beta)x_{T+1} - \sigma(\beta i_{T+1} - \pi_{T+1}) + \sigma r_T^n\big) \tag{A.1}$$

$$\pi_t = \kappa x_t + \hat{\mathbb{E}}_t \sum_{T=t}^{\infty}(\alpha\beta)^{T-t}\big(\kappa\alpha\beta x_{T+1} + (1-\alpha)\beta\pi_{T+1} + u_T\big) \tag{A.2}$$

$$i_t = \psi_\pi \pi_t + \psi_x x_t + \bar{i}_t \qquad \text{(if imposed)} \tag{A.3}$$

$$\text{PLM:} \qquad \hat{\mathbb{E}}_t z_{t+h} = a_{t-1} + b h_x^{h-1} s_t \quad \forall h \geq 1 \qquad b = g_x\, h_x \tag{A.4}$$

$$\text{Updating:} \qquad a_t = a_{t-1} + k_t^{-1}\big(z_t - (a_{t-1} + b s_{t-1})\big) \tag{A.5}$$

$$\text{Anchoring function:} \qquad k_t^{-1} = \rho_k k_{t-1}^{-1} + \gamma_k f e_{t-1}^2 \tag{A.6}$$

$$\text{Forecast error:} \qquad f e_{t-1} = z_t - (a_{t-1} + b s_{t-1}) \tag{A.7}$$

$$\text{LH expectations:} \qquad f_a(t) = \frac{1}{1-\alpha\beta} a_{t-1} + b(\mathbb{I}_{nx} - \alpha\beta h)^{-1} s_t \qquad f_b(t) = \frac{1}{1-\beta} a_{t-1} + b(\mathbb{I}_{nx} - \beta h)^{-1} s_t \tag{A.8}$$

This notation captures vector learning (z learned) for intercept only. For scalar learning, $a_t = \begin{pmatrix}\bar{\pi}_t & 0 & 0\end{pmatrix}'$ and b_1 designates the first row of b. The observables (π, x) are determined as:

$$x_t = -\sigma i_t + \begin{bmatrix}\sigma & 1-\beta & -\sigma\beta\end{bmatrix} f_b + \sigma \begin{bmatrix}1 & 0 & 0\end{bmatrix} (\mathbb{I}_{nx} - \beta h_x)^{-1} s_t \tag{A.9}$$

$$\pi_t = \kappa x_t + \begin{bmatrix}(1-\alpha)\beta & \kappa\alpha\beta & 0\end{bmatrix} f_a + \begin{bmatrix}0 & 0 & 1\end{bmatrix} (\mathbb{I}_{nx} - \alpha\beta h_x)^{-1} s_t \tag{A.10}$$

B Target criterion

The target criterion in the simplified model (scalar learning of inflation intercept only, $k_t^{-1} = \mathbf{g}(f e_{t-1})$):

$$\pi_t = -\frac{\lambda_x}{\kappa}\Bigg\{x_t - \frac{(1-\alpha)\beta}{1-\alpha\beta}\Big(k_t^{-1} + ((\pi_t - \bar{\pi}_{t-1} - b_1 s_{t-1}))\mathbf{g}_\pi(t)\Big)$$
$$\Big(\mathbb{E}_t \sum_{i=1}^{\infty} x_{t+i} \prod_{j=0}^{i-1}(1 - k_{t+1+j}^{-1} - (\pi_{t+1+j} - \bar{\pi}_{t+j} - b_1 s_{t+j})\mathbf{g}_{\bar{\pi}}(t+j))\Big)\Bigg\} \tag{B.1}$$

where I'm using the notation that $\prod_{j=0}^{0} \equiv 1$. For interpretation purposes, let me rewrite this as follows:

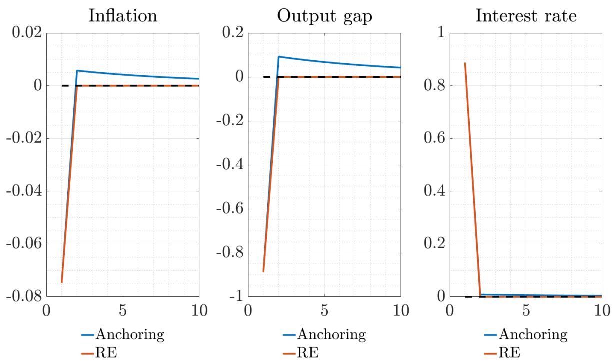
$$\pi_t = -\frac{\lambda_x}{\kappa} x_t + \frac{\lambda_x}{\kappa}\frac{(1-\alpha)\beta}{1-\alpha\beta}\Big(k_t^{-1} + f e_{t|t-1}^{eve}\mathbf{g}_\pi(t)\Big)\mathbb{E}_t \sum_{i=1}^{\infty} x_{t+i}$$
$$-\frac{\lambda_x}{\kappa}\frac{(1-\alpha)\beta}{1-\alpha\beta}\Big(k_t^{-1} + f e_{t|t-1}^{eve}\mathbf{g}_\pi(t)\Big)\Big(\mathbb{E}_t \sum_{i=1}^{\infty} x_{t+i} \prod_{j=0}^{i-1}(k_{t+1+j}^{-1} + f e_{t+1+j|t+j}^{eve})\mathbf{g}_{\bar{\pi}}(t+j)\Big) \tag{B.2}$$

Interpretation: tradeoffs from discretion in RE + effect of current level and change of the gain on future tradeoffs + effect of future expected levels and changes of the gain on future tradeoffs

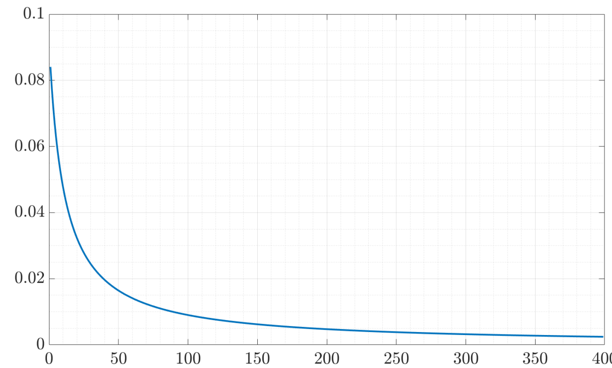
C Impulse responses to iid monpol shocks across a wide range of learning models

$T = 400, N = 100, n_{drop} = 5$, shock imposed at $t = 25$, calibration as above, Taylor rule assumed to be known, PLM = learn constant only, of inflation only.

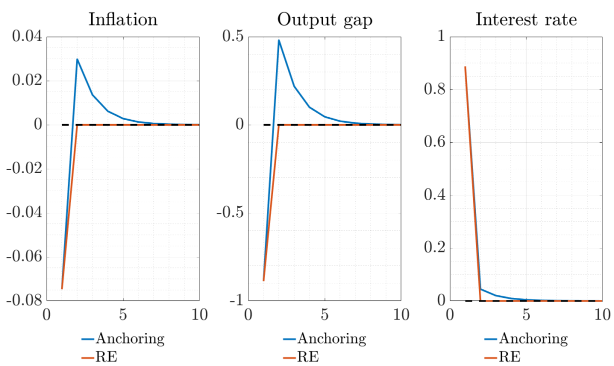
Figure 21: IRFs and gain history (sample means)



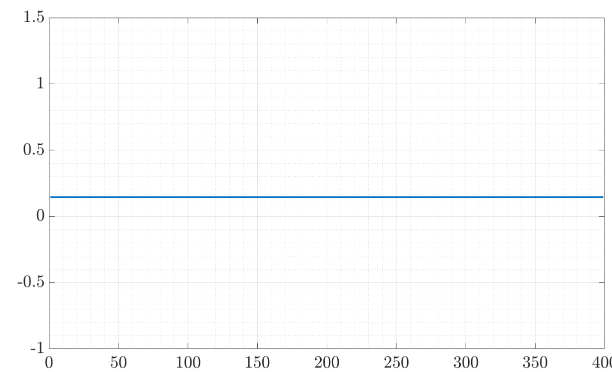
(a) Decreasing gain learning



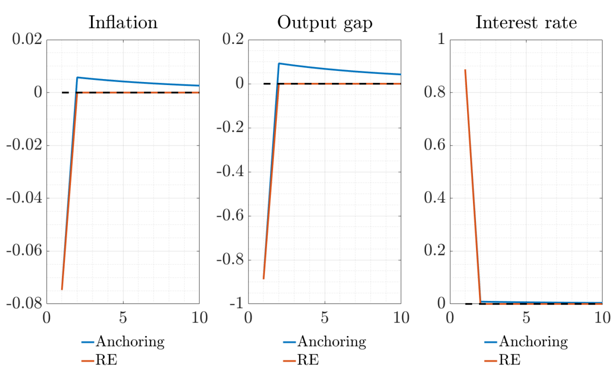
(b) Mean gain



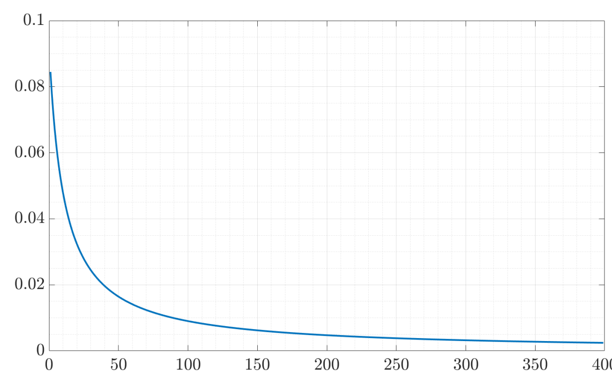
(c) Constant gain learning



(d) Mean gain

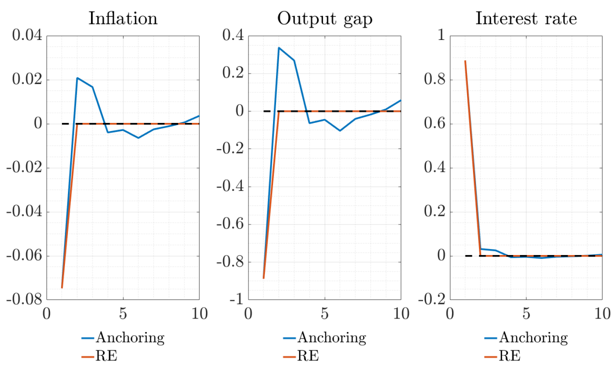


(e) CEMP criterion (vector)

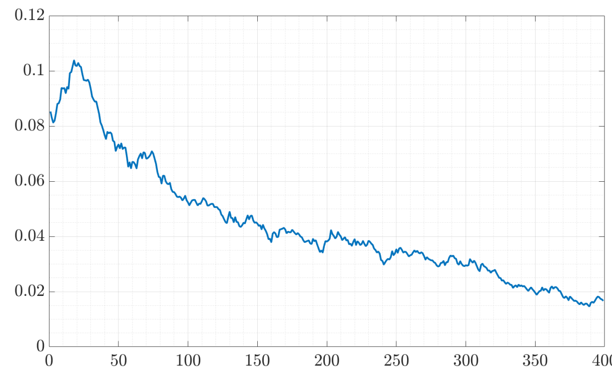


(f) Mean gain

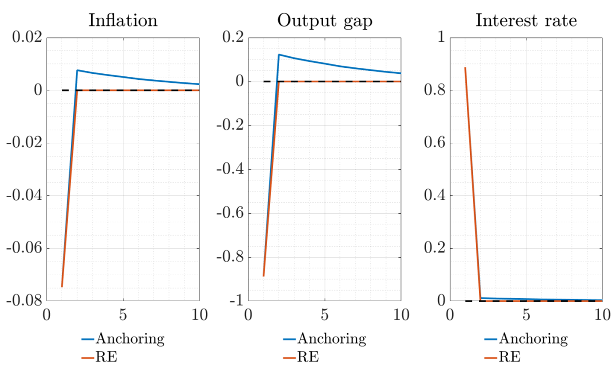
Figure 22: IRFs and gain history (sample means), continued



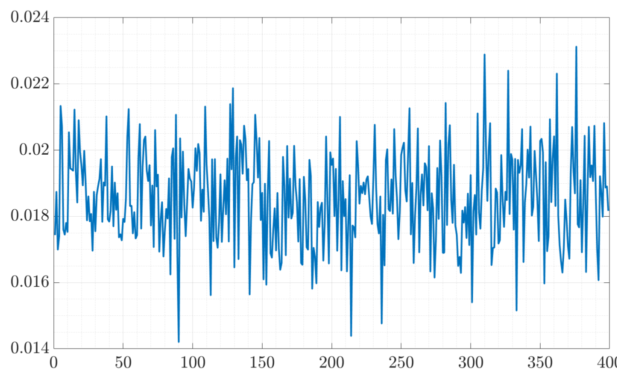
(a) CUSUM criterion (vector)



(b) Mean gain



(c) Smooth criterion, approximated, using $\alpha^{true} = (0.05; 0.025; 0; 0.025; 0.05)$, on $fe \in (-2, 2)$.



(d) Mean gain