Materials 10 - Is overshooting endemic to constant gain learning?

Laura Gáti

November 19, 2019

Overview

1 Model summary

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

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

$$i_t = \psi_\pi \pi_t + \psi_x x_t + \rho i_{t-1} + \bar{i}_t \tag{3}$$

I consider two variations of the learning rule. The first is a "mean-only" rule:

$$\hat{\mathbb{E}}_t z_{t+h} = \begin{bmatrix} \bar{\pi}_{t-1} \\ 0 \\ 0 \end{bmatrix} + b h_x^{h-1} s_t \quad \forall h \geq 1 \quad b = g_x\ h_x, \qquad \text{PLM1} \tag{4}$$

$$\text{but the first row of } b \text{ is} \quad b_1 = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \tag{5}$$

$$\bar{\pi}_t = \bar{\pi}_{t-1} + k_t^{-1} \underbrace{\big(\pi_t - \bar{\pi}_{t-1}\big)}_{\text{fcst error using (4)}} \tag{6}$$

The second is a "learning the slope too" rule:

$$\hat{\mathbb{E}}_t z_{t+h} = \begin{bmatrix} \bar{\pi}_{t-1} \\ 0 \\ 0 \end{bmatrix} + b_{t-1} h_x^{h-1} s_t \quad \forall h \geq 1 \quad b = g_x\ h_x, \qquad \text{PLM2} \tag{7}$$

$$\text{but the first row of } b \text{ is } b_{1,t} \text{ and is also learned. Let } \phi_t = \begin{bmatrix} \bar{\pi}_t & b_{1,t} \end{bmatrix} \tag{8}$$

$$\phi_t = \left(\phi_{t-1}' + k_t^{-1} \underbrace{\left(\pi_t - \phi_{t-1} \begin{bmatrix} 1 \\ s_{t-1} \end{bmatrix} \right)}_{\text{fcst error using (7)}} \right)' \tag{9}$$

2 Compact notation

$$z_t = A_p^{RE}\,\mathbb{E}_t\, z_{t+1} + A_s^{RE} s_t \tag{10}$$

$$z_t = A_a^{LH} f_a(t) + A_b^{LH} f_b(t) + A_s^{LH} s_t \tag{11}$$

$$s_t = P s_{t-1} + \epsilon_t \qquad \rightarrow \quad s'_t = hx\; s'_{t-1} + \epsilon'_t \tag{12}$$

$$\text{where} \quad s'_t \equiv \begin{pmatrix} r_t^n \\ \bar{i}_t \\ u_t \\ i_{t-1} \end{pmatrix} \quad hx \equiv \begin{pmatrix} \rho_r & 0 & 0 & 0 \\ 0 & \rho_i & 0 & 0 \\ 0 & 0 & \rho_u & 0 \\ gx_{3,1} & gx_{3,2} & gx_{3,3} & gx_{3,4} \end{pmatrix} \quad \epsilon'_t \equiv \begin{pmatrix} \varepsilon_t^r \\ \varepsilon_t^i \\ \varepsilon_t^u \\ 0 \end{pmatrix} \quad \text{and} \quad \Sigma' = \begin{pmatrix} \sigma_r & 0 & 0 & 0 \\ 0 & \sigma_i & 0 & 0 \\ 0 & 0 & \sigma_u & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{13}$$

And the A_s^{RE} and A_s^{LH} are given by:

$$A_s^{RE} = \begin{pmatrix} \frac{\kappa\sigma}{w} & -\frac{\kappa\sigma}{w} & 1 - \frac{\kappa\sigma\psi_\pi}{w} & 0 \\ \frac{\sigma}{w} & -\frac{\sigma}{w} & -\frac{\sigma\psi_\pi}{w} & 0 \\ \psi_x(\frac{\sigma}{w}) + \psi_\pi(\frac{\kappa\sigma}{w}) & \psi_x(-\frac{\sigma}{w}) + \psi_\pi(-\frac{\kappa\sigma}{w}) + 1 & \psi_x(-\frac{\sigma\psi_\pi}{w}) + \psi_\pi(1 - \frac{\kappa\sigma\psi_\pi}{w}) & \rho \end{pmatrix} \tag{14}$$

$$A_s^{LH} = \begin{pmatrix} g_{\pi s} \\ g_{xs} \\ \psi_\pi g_{\pi s} + \psi_x g_{xs} + \begin{bmatrix} 0 & 1 & 0 & \rho \end{bmatrix} \end{pmatrix} \tag{15}$$

$$g_{\pi s} = (1 - \frac{\kappa\sigma\psi_\pi}{w}) \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} (I_4 - \alpha\beta hx)^{-1} - \frac{\kappa\sigma}{w} \begin{bmatrix} -1 & 1 & 0 & \rho \end{bmatrix} (I_4 - \beta hx)^{-1} \tag{16}$$

$$g_{xs} = \frac{-\sigma\psi_\pi}{w} \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} (I_4 - \alpha\beta hx)^{-1} - \frac{\sigma}{w} \begin{bmatrix} -1 & 1 & 0 & \rho \end{bmatrix} (I_4 - \beta hx)^{-1} \tag{17}$$

And long-horizon expectations are

$$f_a(t) \equiv \hat{\mathbb{E}}_t \sum_{T=t}^{\infty} (\alpha\beta)^{T-t} \begin{bmatrix} \pi_{T+1} \\ x_{T+1} \\ i_{T+1} \end{bmatrix} \qquad f_b(t) \equiv \hat{\mathbb{E}}_t \sum_{T=t}^{\infty} (\beta)^{T-t} \begin{bmatrix} \pi_{T+1} \\ x_{T+1} \\ i_{T+1} \end{bmatrix} \tag{18}$$

$$f_a(t) = \frac{1}{1-\alpha\beta} \begin{bmatrix} \bar{\pi}_t \\ 0 \\ 0 \end{bmatrix} + b(I_4 - \alpha\beta h_x)^{-1} s_t \qquad f_b(t) = \frac{1}{1-\beta} \begin{bmatrix} \bar{\pi}_t \\ 0 \\ 0 \end{bmatrix} + b(I_4 - \beta h_x)^{-1} s_t \tag{19}$$

3 Recap of timing

Define some objects: *(I usually let t denote the time in which the variable is formed.)*

$$f_t^j = \hat{\mathbb{E}}_t(z_{t+1}) \quad \text{one-period-ahead forecast formed at time } t,\ j = m, e \text{ (morning or evening)} \tag{20}$$

$$FE_t = z_{t+1} - f_t \quad \text{one-period-ahead forecast error realized at time } t+1 \tag{21}$$

$$= ALM(t+1) - PLM(t) \tag{22}$$

$$\theta_t = \hat{\mathbb{E}}_{t-1}(z_t) - \mathbb{E}_{t-1}(z_t) \quad \text{CEMP's criterion} \tag{23}$$

$$= PLM(t-1) - \mathbb{E}_{t-1}\, ALM(t) \tag{24}$$

$PLM(t) : \hat{\mathbb{E}}_t z_{t+1} = \bar{z}_{t-1} + bs_t$

Morning: morning of time t available: $\mathcal{I}_t^m = \{\bar{z}_{t-1}, s_t, k_{t-1}, FE_{t-2}\}$

1. Form all future expectations using $PLM(t)$(morning forecast) $\rightarrow z_t$ realized, $\rightarrow FE_{t-1}$ realized

2. Form $\theta_t \rightarrow k_t$ realized

3. **Evening**: Update $\bar{z}_t = \bar{z}_{t-1} + k_t^{-1}(FE_{t-1}^e)$

   where $FE_{t-1}^e = z_t - f_{t-1}^e = z_t - (\bar{z}_{t-1} + bs_{t-1})$ is the most recent realized FE, so:

   $\bar{z}_t = \bar{z}_{t-1} + k_t^{-1}(z_t - (\bar{z}_{t-1} + bs_{t-1}))$

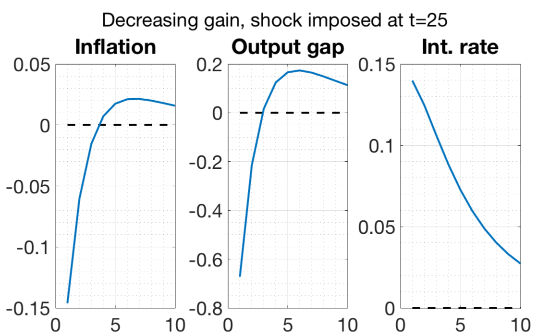
   $\rightarrow$ evening of time t available: $\mathcal{I}_t^e = \{\bar{z}_t, s_t, k_t, FE_{t-1}\}$

4 Current set of baseline parameters

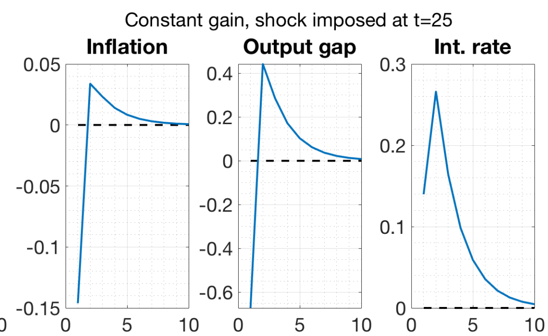
β	0.99	stochastic discount factor	standard (Woodford 2003/2011)
σ	1	IES	consistent with balanced growth
α	0.5	Calvo probability of not adjusting	match 6-month duration of prices (can increase to 0.75)
ψ_π	1.5	coefficient of inflation in Taylor rule	Taylor
ψ_x	0	coefficient of output gap in Taylor rule	focus on π
$\bar{g}$	0.145	value of the constant gain	CEMP
$\bar{\theta}$	1	threshold deviation between $\hat{\mathbb{E}}$ & $\mathbb{E}$	CEMP: 0.029
ρ_r	0	persistence of natural rate shock	n.a.
ρ_i	0.6	persistence of monetary policy shock	CEMP: 0.877 (can increase to 0.78 if $\alpha = 0.75$)
ρ_u	0	persistence of cost-push shock	CEMP
σ_r	0.1	standard deviation of natural rate shock	n.a.
σ_i	0.359	standard deviation of mon. policy shock	CEMP
σ_u	0.277	standard deviation of cost-push shock	CEMP
θ	10	price elasticity of demand	Woodford 2003/2011, Chari, Kehoe & McGrattan 2000
ω	1.25	elasticity of marginal cost to output	Woodford 2003/2011, Chari, Kehoe & McGrattan 2000

5 Cross-sectional IRFs, mon. pol shock only, cgain & dgain only, “mean-only” PLM ◀

Figure 1: IRF for observables, shock imposed at t

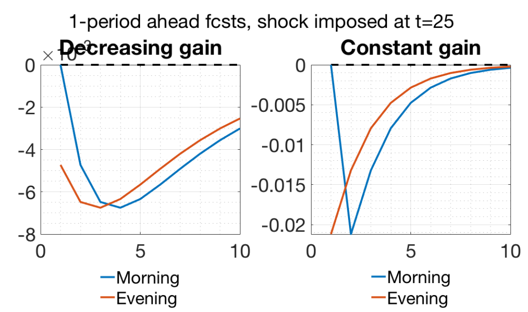


(a) Decreasing gain, $t = 25$

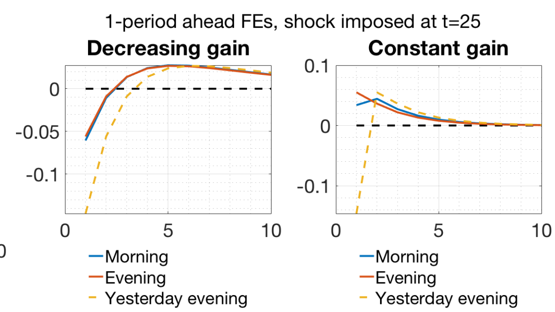


(b) Constant gain, $t = 25$

Figure 2: IRF for 1-period ahead forecasts and FEs, together, morning and evening, shock imposed at t

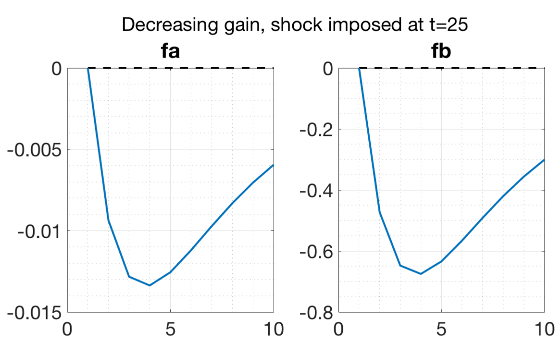


(a) Forecasts, $t = 25$

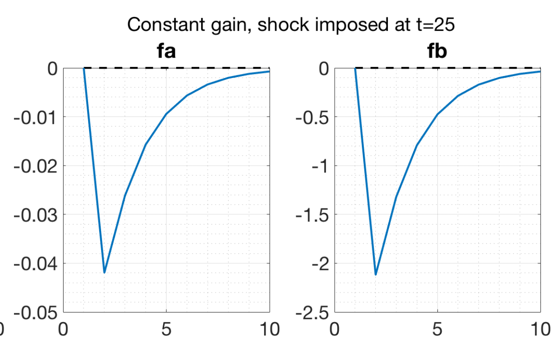


(b) Forecast errors, $t = 25$

Figure 3: IRF for LH forecasts, shock imposed at t



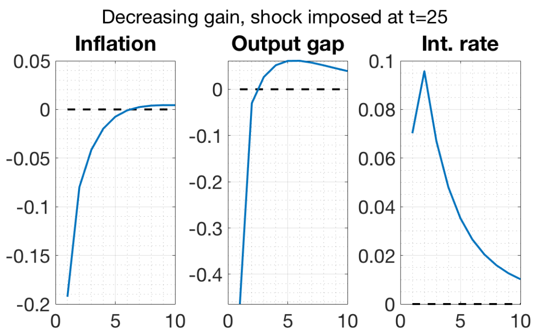
(a) Decreasing gain, $t = 25$



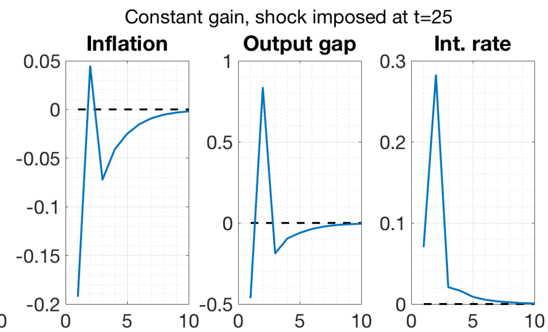
(b) Constant gain, $t = 25$

6 Cross-sectional IRFs, mon. pol shock only, cgain & dgain only, "slope and constant" PLM ◄

Figure 4: IRF for observables, shock imposed at t

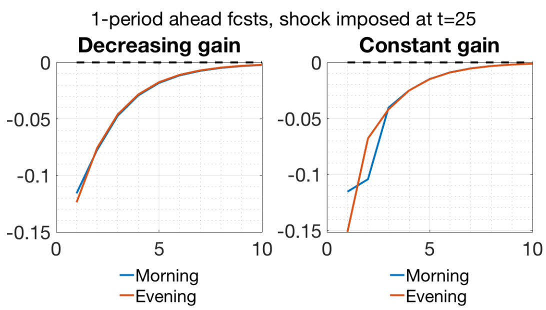


(a) Decreasing gain, $t = 25$

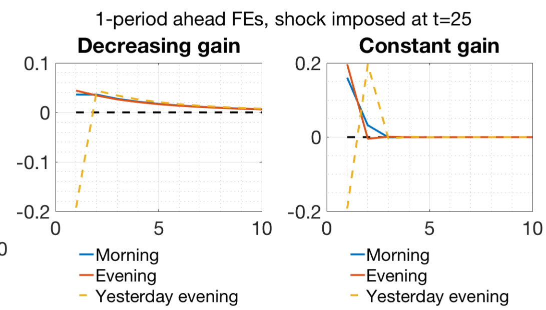


(b) Constant gain, $t = 25$

Figure 5: IRF for 1-period ahead forecasts and FEs, together, morning and evening, shock imposed at t

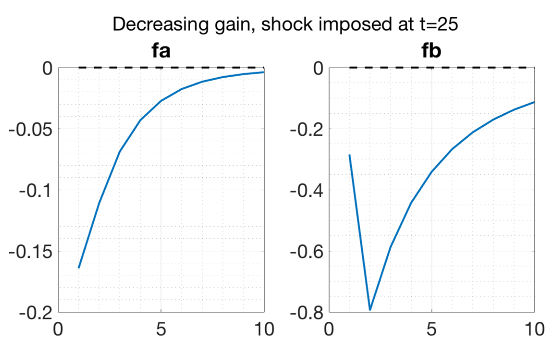


(a) Forecasts, $t = 25$

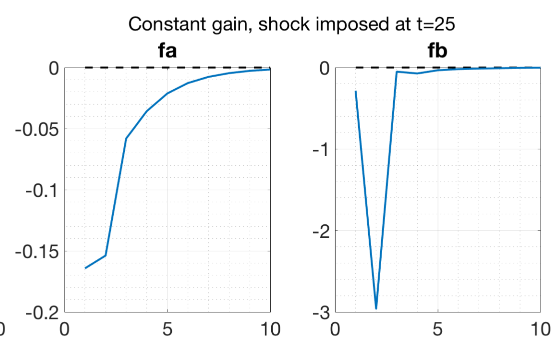


(b) Forecast errors, $t = 25$

Figure 6: IRF for LH forecasts, shock imposed at t



(a) Decreasing gain, $t = 25$



(b) Constant gain, $t = 25$

- is almost identical to constant-only learning b/c 1) they're only learning the slope of inflation 2) f_a, f_b are still driven mainly by the constant.

7 Choosing $\bar{g}$ to minimize the FEV ◀

Currently what I do is:

- For each simulated sequence of shocks n
- calculate the FEV across time for that particular history as a function of $\bar{g}$
- choose $\bar{g}_n^*$ to minimize the FEV
- take an average across the simulations

For 500 simulated sequences, I obtain an average $\bar{g}^* = 2.5715 \times 10^{-4} \approx 0.0003$. The maximum value for $\bar{g}_n^*$ is 0.0049, and I constrain $\bar{g}_n^*$ to lie between $[0.00001, 0.2]$. (Somewhat troubling is that without the constraint, $\bar{g}_n^*$ is often negative, although small.)

With a $\bar{g}^* = 0.0003$, the overshooting is completely killed and decreasing and constant gain learning look identical. FEs still switch sign one time, but barely because, after impact, they are extremely small.

A note: Ideally, I wanted to take the FEV over the cross-section instead of across time. This is however much more computationally intensive because it requires for each history n and each time period t to resimulate all sequences $1, \dots, N$ up to time t to compute the most recent, period t FEV across the cross-section, and then for every proposed value of $\bar{g}$, to repeat this process until $\bar{g}_t^*$ is found. Thus I expect this to take at least T times longer than the first approach (and likely more because fmincon will also take at least N times longer). Considering that the "across-time" approach takes a little above 5 minutes to run, this second, "cross-section" approach would need *at least* 50 minutes for a modest simulation length of 100.

8 How observables respond to expectations - connecting RE and learning

For this section, disregard the difference in the expectation operator in the two models. Pretend like it was the same operator. With this assumption, the RE model is just a recursive formulation of the learning model. My aim here is to show that both ways of writing the *same* system embody the same channels of how observables respond to expectations, only these channels are more explicit in the non-recursive formulation.

Ignoring shocks and setting $\psi_x = 0$, so the Taylor rule is just $i_t = \psi_\pi \pi_t$, the two systems are

RE

$$x_t = -\sigma i_t + \mathbb{E}_t\, x_{t+1} + \sigma\, \mathbb{E}_t\, \pi_{t+1}$$

$$\pi_t = \kappa x_t + \beta\, \mathbb{E}_t\, \pi_{t+1}$$

Learning

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

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

When you plug in the interest rate, you see that the recursive representation hides the contractionary effect of future positive values on current x_t (and thereby on inflation) coming from the anticipated interest rate effect. (Throughout I'm using blue to denote negative values.)

RE

$$x_t = -\sigma\psi_\pi \pi_t + \mathbb{E}_t\, x_{t+1} + \sigma\, \mathbb{E}_t\, \pi_{t+1}$$

$$\pi_t = \kappa x_t + \beta\, \mathbb{E}_t\, \pi_{t+1}$$

Learning

$$x_t = -\sigma\psi_\pi \pi_t + \hat{\mathbb{E}}_t \sum_{T=t}^{\infty} \beta^{T-t}\big((1-\beta)x_{T+1} + \sigma(1-\beta\psi_\pi)\pi_{T+1}\big)$$

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

This gives

RE

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

$$\pi_t = \overbrace{\left(\frac{\kappa\sigma(1-\beta\psi_\pi)}{1+\sigma\psi_\pi\kappa} + \beta\right)}^{+}\mathbb{E}_t\,\pi_{t+1} + \frac{\kappa}{1+\sigma\psi_\pi\kappa}\,\mathbb{E}_t\,x_{t+1}$$

Learning

$$x_t = \left(-\frac{\sigma\psi_\pi}{1+\sigma\psi_\pi\kappa}(1-\alpha)\beta + \frac{\sigma(1-\beta\psi_\pi)}{1+\sigma\psi_\pi\kappa}\right)\mathbb{E}_t^{\alpha,\beta}\,\pi_\infty + \overbrace{\left(-\frac{\sigma\psi_\pi}{1+\sigma\psi_\pi\kappa}\kappa\alpha\beta + \frac{1-\beta}{1+\sigma\psi_\pi\kappa}\right)}^{-}\mathbb{E}_t^{\alpha,\beta}\,x_\infty$$

$$\pi_t = \overbrace{\left((1-\frac{\kappa\sigma\psi_\pi}{1+\sigma\psi_\pi\kappa})(1-\alpha)\beta + \frac{\kappa\sigma(1-\beta\psi_\pi)}{1+\sigma\psi_\pi\kappa}\right)}^{+}\mathbb{E}_t^{\alpha,\beta}\,\pi_\infty + \left((1-\frac{\kappa\sigma\psi_\pi}{1+\sigma\psi_\pi\kappa})\kappa\alpha\beta + \frac{\kappa(1-\beta)}{1+\sigma\psi_\pi\kappa}\right)\mathbb{E}_t^{\alpha,\beta}\,x_\infty$$

where I write $\mathbb{E}_t^{\alpha,\beta}\,x_\infty$ for $\mathbb{E}_t\sum_{T=t}^{\infty}(\alpha\beta)^{T-t}x_{T+1} + \mathbb{E}_t\sum_{T=t}^{\infty}(\beta)^{T-t}x_{T+1}$

8.1 A more concise rephrasing ◀

Ignoring shocks and setting $\psi_x = 0$, so the Taylor rule is just $i_t = \psi_\pi \pi_t$, the two systems are (throughout I'm using blue to denote negative values).

RE

$$x_t = -\sigma\psi_\pi \pi_t + \mathbb{E}_t\, x_{t+1} + \sigma\, \mathbb{E}_t\, \pi_{t+1}$$

$$\pi_t = \kappa x_t + \beta\, \mathbb{E}_t\, \pi_{t+1}$$

Learning

$$x_t = -\sigma\psi_\pi \pi_t + \hat{\mathbb{E}}_t \sum_{T=t}^{\infty} \beta^{T-t}\big((1-\beta)x_{T+1} + \sigma(1-\beta\psi_\pi)\pi_{T+1}\big)$$

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

Expressing x, π as functions of expectations alone, this gives:

RE

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

$$\pi_t = \overbrace{\left(\frac{\kappa\sigma(1-\beta\psi_\pi)}{1+\sigma\psi_\pi\kappa} + \beta\right)}^{+} \mathbb{E}_t\, \pi_{t+1} + \frac{\kappa}{1+\sigma\psi_\pi\kappa}\, \mathbb{E}_t\, x_{t+1}$$

Learning

$$x_t = \frac{-\sigma\psi_\pi}{w} \begin{bmatrix} (1-\alpha)\beta & \kappa\alpha\beta & 0 \end{bmatrix} f_a + \frac{1}{w} \begin{bmatrix} \sigma(1-\beta\psi_\pi) & 1-\beta & 0 \end{bmatrix} f_b$$

$$\pi_t = (1 - \frac{\kappa\sigma\psi_\pi}{w}) \begin{bmatrix} (1-\alpha)\beta & \kappa\alpha\beta & 0 \end{bmatrix} f_a + \frac{\kappa}{w} \begin{bmatrix} \sigma(1-\beta\psi_\pi) & 1-\beta & 0 \end{bmatrix} f_b$$

This yields the stylized representation of how endogenous variables respond to expectations in the two formulations:

RE

$$x_t = \overset{-}{\mathbb{E}(\pi)} + \overset{+}{\mathbb{E}(x)}$$

$$\pi_t = \overset{+}{\mathbb{E}(\pi)} + \overset{+}{\mathbb{E}(x)}$$

Learning

$$x_t = \overset{-}{\mathbb{E}_a(\pi)} + \overset{-}{\mathbb{E}_b(\pi)} + \overbrace{\overset{-}{\mathbb{E}_a(x)} + \overset{+}{\mathbb{E}_b(x)}}^{+ \text{ since } f_a < f_b}$$

$$\pi_t = \underbrace{\overset{+}{\mathbb{E}_a(\pi)} + \overset{-}{\mathbb{E}_b(\pi)}}_{+ \text{ since } \kappa \text{ tiny}} + \overset{+}{\mathbb{E}_a(x)} + \overset{+}{\mathbb{E}_b(x)}$$

As a reminder, I restate the definition of the long-horizon expectations f_a and f_b from Equations

(18) and (19):

$$f_a(t) \equiv \hat{\mathbb{E}}_t \sum_{T=t}^{\infty} (\alpha\beta)^{T-t} \begin{bmatrix} \pi_{T+1} \\ x_{T+1} \\ i_{T+1} \end{bmatrix} \qquad f_b(t) \equiv \hat{\mathbb{E}}_t \sum_{T=t}^{\infty} (\beta)^{T-t} \begin{bmatrix} \pi_{T+1} \\ x_{T+1} \\ i_{T+1} \end{bmatrix} \tag{25}$$

$$f_a(t) = \frac{1}{1-\alpha\beta} \begin{bmatrix} \bar{\pi}_t \\ 0 \\ 0 \end{bmatrix} + b(I_4 - \alpha\beta h_x)^{-1} s_t \qquad f_b(t) = \frac{1}{1-\beta} \begin{bmatrix} \bar{\pi}_t \\ 0 \\ 0 \end{bmatrix} + b(I_4 - \beta h_x)^{-1} s_t \tag{26}$$

(And $b = g_x h_x$, where h_x is the state transition matrix and g_x is the observation matrix from the RE model solution.)

Looking at (26), it's obvious that a) learning will show up in the intercept, as the second part will load the same on the shocks in each period; b) because of much heavier discounting of the intercept, f_a will fluctuate a lot less than f_b. Therefore the effects coming from f_b overweigh those from f_a, which is why the blue difference in the stylized representation goes away.