# Materials 15 - More on the CEMP vs. CUSUM criteria and optimal Taylor rule coefficients

Laura Gáti

February 4, 2020

## Overview

# 1  Model summary

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

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

$$i_t = \psi_\pi \pi_t + \psi_x x_t + \bar{i}_t \tag{3}$$

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

$$\bar{z}_t = \bar{z}_{t-1} + k_t^{-1} \underbrace{\big(z_t - (\bar{z}_{t-1} + b s_{t-1})\big)}_{\text{fcst error using (4)}} \tag{5}$$

(Vector learning. For scalar learning, $\bar{z} = \begin{pmatrix} \bar{\pi} & 0 & 0 \end{pmatrix}'$. I'm also not writing the case where the slope $b$ is also learned.)

$$k_t = \begin{cases} k_{t-1} + 1 & \text{for decreasing gain learning} \\[2ex] \bar{g}^{-1} & \text{for constant gain learning.} \end{cases} \tag{6}$$

**Figure 1:** Reference: baseline model



(a) Learning constant only  (b) Learning slope and constant
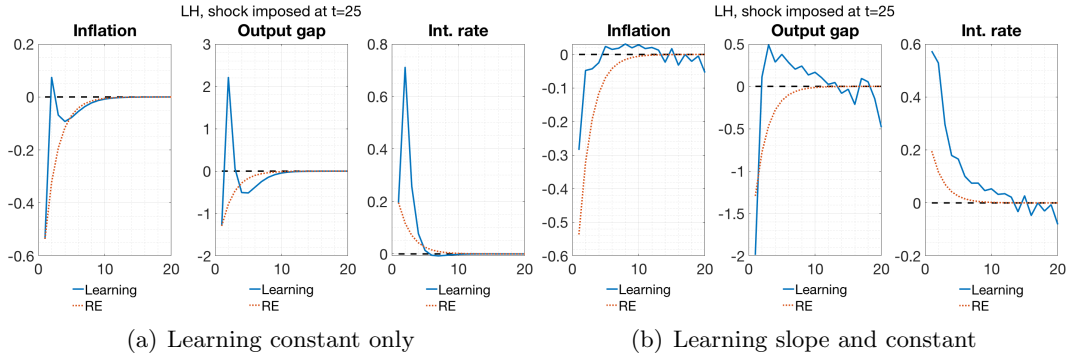
## 2 The CEMP vs. the CUSUM criterion

CEMP's criterion

$$\theta_t = |\hat{\mathbb{E}}_{t-1}\pi_t - \mathbb{E}_{t-1}\,\pi_t|/(\text{Var}(\text{shocks})) \tag{7}$$

$$\text{i.e.} \quad \text{PLM- E[ALM], scaled by shocks} \tag{8}$$

For my version of CEMP's criterion, I rewrite the ALM

$$z_t = A_a f_a + A_b f_b + A_s s_t \tag{9}$$

$$\text{as} \quad z_t = F + G s_t \tag{10}$$

$$\Leftrightarrow \quad z_t = \begin{bmatrix} F & G \end{bmatrix} \begin{bmatrix} 1 \\ s_t \end{bmatrix} \tag{11}$$

Then, since the PLM is $z_t = \phi \begin{bmatrix} 1 \\ s_t \end{bmatrix}$, the generalized CEMP criterion becomes

$$\theta_t = \max |\Sigma^{-1}(\phi - \begin{bmatrix} F & G \end{bmatrix})| \tag{12}$$

where $\Sigma$ is the VC matrix of shocks. As for the CUSUM criterion, what I did in Materials 5 was

$$\omega_t = \omega_{t-1} + \kappa k_{t-1}^{-1}(FE_t^2 - \omega_{t-1}) \tag{13}$$

$$\theta_t = \theta_{t-1} + \kappa k_{t-1}^{-1}(FE_t^2/\omega_t - \theta_{t-1}) \tag{14}$$

where $FE_t$ is the most recent short-run forecast error ($ny \times 1$), and $\omega_t$ is the agents' estimate of the forecast error variance ($ny \times ny$). To take into account that these are now matrices, I now write

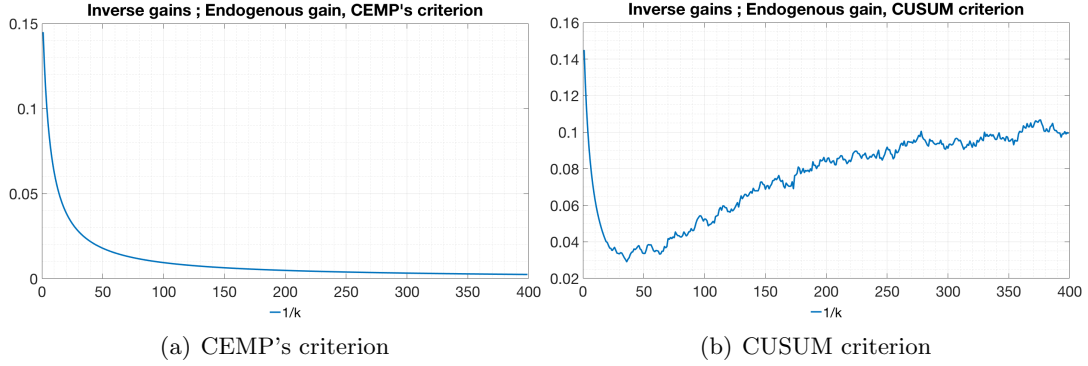$$\omega_t = \omega_{t-1} + \kappa k_{t-1}^{-1}(f_t f_t' - \omega_{t-1}) \tag{15}$$

$$\theta_t = \theta_{t-1} + \kappa k_{t-1}^{-1}(f_t' \omega_t^{-1} f_t - \theta_{t-1}) \tag{16}$$

(Note: I'm using Lütkepohl's *Introduction to Multiple Time Series Analysis*, p. 160 to reformulate the CUSUM criterion as a statistic that has a $\chi^2$ distribution.)

# 3 Investigating the behavior of CEMP and CUSUM criteria

## 3.1 Anchoring as a function of $\psi_\pi$, fixing $\psi_x = 0, \bar{\theta} = 4, \tilde{\theta} = 2.5$

**Figure 2:** Inverse gains, $\psi_\pi = 1.01$
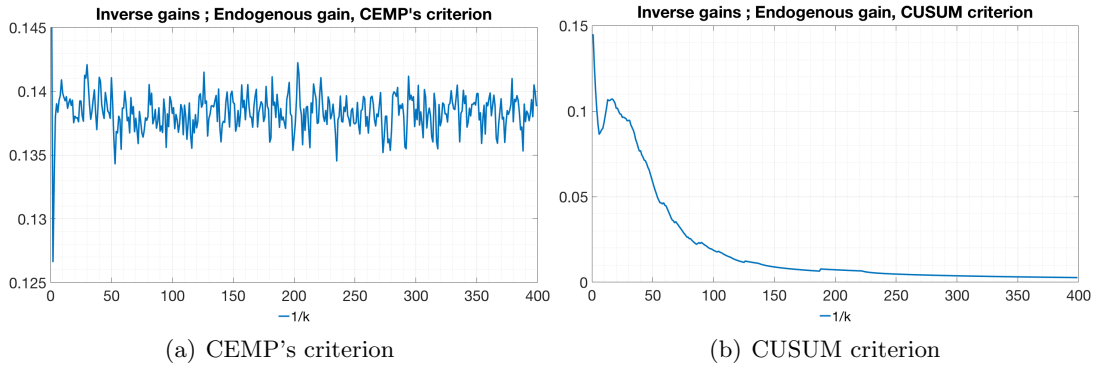


(a) CEMP's criterion

(b) CUSUM criterion

**Figure 3:** Inverse gains, $\psi_\pi = 1.5$



(a) CEMP's criterion

(b) CUSUM criterion

**Figure 4:** Inverse gains, $\psi_\pi = 2$



(a) CEMP's criterion

(b) CUSUM criterion
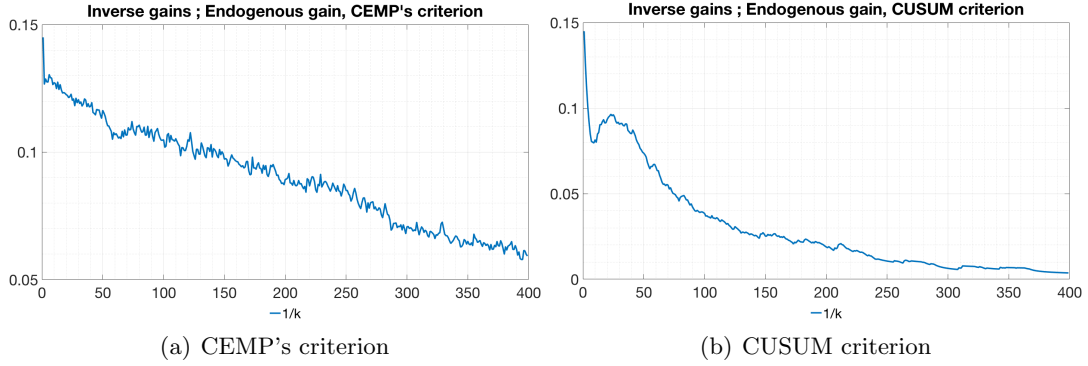
## 3.2  Why do the two criteria behave opposite ways?

A restatement of the two criteria: you get unanchored expectations if:

$$\theta_t^{CEMP} = \max(\Sigma^{-1}|(\phi - \begin{bmatrix} F & G \end{bmatrix})|) > \bar{\theta} \qquad \text{vs.} \qquad \theta_t^{CUSUM} = f'\omega^{-1}f > \tilde{\theta} \tag{17}$$

where $\phi$ is the agents' estimated matrix, $F, G$ are the ALM matrices, $f$ is the one-period ahead forecast error and $\omega$ is the estimated forecast error variance matrix.

Sloppily, we can write the two criteria as:

$$\frac{\mathbb{E}(FE)}{\text{Var(noise)}} \qquad \text{vs.} \qquad \frac{FE^2}{F\hat{E}V} \tag{18}$$

- The big difference in the two is the way they treat noise: CEMP integrates it out, while CUSUM tries to cancel it out by dividing by the estimated FEV.

- Squaring the numerator in the CEMP criterion doesn't change its behavior.

- Trying various ways of not squaring or taking the square root of the CUSUM criterion doesn't change its behavior.

- Inputting Var(noise) instead of $F\hat{E}V$ in the CUSUM denominator changes its behavior! You obtain the CEMP-criterion behavior: little anchoring for high $\psi_\pi$, much anchoring for low $\psi_\pi$.

$\rightarrow$ Is the difference due to using the $FEV$ instead of Var(noise) or does it come from using an *estimated* FEV?

- Here it is: Can decompose forecast errors as $FE = $ part due to learning $+$ noise. CEMP's criterion is only moved by the former, while the CUSUM criterion moves in response to both. So imagine now a situation that introduces errors due to learning (high $\psi_\pi$): CEMP's criterion goes up. But CUSUM-agents can't distinguish between the two types of errors, so not only does FE go up, but also the estimate of the FEV. So the criterion doesn't move as much.

*A small side-note:*

- *the CUSUM-criterion is very sensitive to initialization:*
  *Especially the choice of $\theta_0$ matters a lot for initial dynamics which is why on these graphs you initially have anchoring b/c I initialize $\theta_0$ as zero. If, on the contrary, I initialize it as $\tilde{\theta}$ or higher, you don't get the initial anchored period. The same effect is achieved by setting a burn-in period of 100 or so.*

# 4 Analytical expressions for optimal Taylor rule coefficients

Following Woodford's *Interest and Prices*, here's a procedure to obtain optimal Taylor rule coefficients. In Woodford's terminology, this consists of solving for the *optimal noninertial plan* (oni) for the endogenous variables, and then doing coefficient comparison between the Taylor rule and the *oni*.

## 4.1 In-a-nutshell algorithm for optimal Taylor rule coefficients

1. Postulate conjectures $z_t = \bar{z} + f_j u_t + g_j \hat{r}_t^n$, $\quad j = \pi, x, i \quad z = (\pi, x, i)'$ for the model consisting of an NKPC and NKIS relation and the AR(1) shocks $u, \hat{r}^n$, where $\hat{r}_t^n = r_t^n - \bar{r}$ (so that the natural rate has a drift, but $\hat{r}^n$ is detrended).

2. Plug the conjectures into the two model equations to derive 2 constraints on the 3 deterministic components $(\bar{\pi}, \bar{x}, \bar{i})$ and 4 constraints on the 6 coefficients on disturbances, $f_j, g_j$, $\quad j = \pi, x, i$. (Use the known LOMs of shocks to write everything in terms of time $t$ shocks.)

3. Solve 3 sets of optimizations

   (a) $(\bar{\pi}, \bar{x}, \bar{i}) = \arg\min L^{det}$ $\quad s.t.$ $\quad$ the 2 constraints on deterministic components

   (b) $f_j = \arg\min L^{stab,u}$ $\quad s.t.$ $\quad$ 2 out of 4 constraints on shock-coefficients, $j = \pi, x, i$.

   (c) $g_j = \arg\min L^{stab,r}$ $\quad s.t.$ $\quad$ the last 2 out of 4 constraints on shock-coefficients, $j = \pi, x, i$.

4. Compare coefficients of Taylor rule to *oni*-solution of $i_t$.

## 4.2 Details

1. The optimal noninertial plan (oni)

   A purely forward-looking set of optimal policies that specifies a LOM for each endogenous variable as the sum of a deterministic component (a long-run average, denoted above by "bar") and a state-contingent component with an optimal response to state $t$ disturbances (the $f_j$ and $g_j$ above). Moreover, the deterministic components are optimal from a timeless perspective (i.e. they minimize $L^{det}$), and the state-contingent components minimize fluctuations coming from shocks ($L^{stab,u}, L^{stab,r}$).

2. The loss function of the monetary authority and its decomposition into deterministic and shock-contingent parts

Following Woodford, I augment my loss function with some concern for interest rate stabilization:

$$L^{CB} = \mathbb{E}_t \sum_{T=t}^{\infty} \{\pi_T^2 + \lambda_x(x_T - x^*)^2 + \lambda_i(i_T - i^*)\} \tag{19}$$

Woodford decomposes this as $L^{CB} = L^{det} + L^{stab}$ where the former only depends on the "deterministic component of the equilibrium paths of the target variables," while the latter "depends only on the equilibrium responses to unexpected shocks" (p. 509). In particular:

$$L^{det} = \sum_{T=t}^{\infty} \beta^{T-t} \{\mathbb{E}_t \pi_T^2 + \lambda_x(\mathbb{E}_t x_T - x^*)^2 + \lambda_i(\mathbb{E}_t i_T - i^*)^2\} \tag{20}$$

$$L^{stab} = \sum_{T=t}^{\infty} \beta^{T-t} \{\text{var}_t(\pi_T) + \lambda_x \text{var}_t(x_T) + \lambda_i \text{var}_t(i_T)\} \tag{21}$$

Woodford then further decomposes $L^{stab}$ into an element conditional on each shock, but this is only for algebraic convenience.

3. The Taylor rule

Woodford postulates a Taylor rule of the form

$$i_t = \bar{i} + \phi_\pi(\pi_t - \bar{\pi}) + \phi_x(x_t - \bar{x})/4 \tag{22}$$

(He divides by 4 to make the output gap quarterly.) Substituting in the conjectured and solved-for *oni*-solutions for the endogenous variables, one obtains:

$$i_t = \bar{i} + \phi_\pi(f_\pi u_t + g_\pi \hat{r}_t^n) + \phi_x(f_x u_t + g_x \hat{r}_t^n)/4 \tag{23}$$

$$i_t = \bar{i} + f_i u_t + g_i \hat{r}_t^n \tag{24}$$

allowing one to solve for $(\phi_\pi^*, \phi_x^*)$ as the solution to

$$f_i = \phi_\pi f_\pi + \phi_x f_x \tag{25}$$

$$g_i = \phi_\pi g_\pi + \phi_x g_x \tag{26}$$

## 4.3 Optimal Taylor rule coefficients for RE model, with the simplifying assumption $\rho_u = \rho_r = \rho$

$$\phi_\pi^* = \frac{\kappa\sigma}{\lambda_i(\rho-1)(\beta\rho-1) - \kappa\lambda_i\rho\sigma} \tag{27}$$

$$\phi_x^* = \frac{\lambda_x\sigma(1-\beta\rho)}{\lambda_i(\rho-1)(\beta\rho-1) - \kappa\lambda_x\rho\sigma} \tag{28}$$

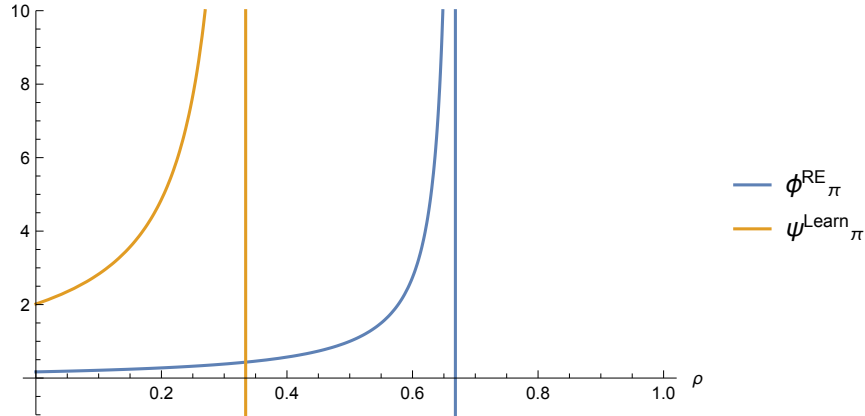which is - fabulously enough - exactly what Woodford obtains.

## 4.4 Optimal Taylor rule coefficients for the learning model, with the simplifying assumption $\rho_u = \rho_r = \rho$

$$\psi_\pi^* = \frac{\kappa\sigma(\beta(\rho-1)-1)(\alpha\beta(\rho-1)-1)}{\kappa\lambda_i\sigma(\alpha\beta(\rho-1)-1) + \beta\lambda_i(\rho-1)(\alpha\beta(\rho-1)+\beta-1)} \tag{29}$$

$$\psi_x^* = \frac{\lambda_x\sigma(\beta(\rho-1)-1)(\alpha\beta(\rho-1)+\beta-1)}{\kappa\lambda_i\sigma(\alpha\beta(\rho-1)-1) + \beta\lambda_i(\rho-1)(\alpha\beta(\rho-1)+\beta-1)} \tag{30}$$

For a simple calibration like my model of $\lambda_i = 1$ and $\lambda_x = 0$ and $\{\beta \to 0.99, \sigma \to 1, \kappa \to 0.16, \rho \to 0.3, \alpha \to 0.5\}$, I get $\phi_\pi^* = 0.360279$ and $\psi_\pi^* = 18.7714$.

**Figure 5:** Optimal Taylor rule coefficients for $\lambda_i = 1$ as a function of $\rho$



Like Woodford observes, equations (27)-(30) impose certain bounds on $\rho$: this bound is higher for RE than for learning. Another point to note is that decreasing $\lambda_i$ just scales the optimal coefficients as a function of $\rho$ up, but preserving the functional form.