Learning in Macroeconomic Models

Wouter J. Den Haan
London School of Economics

Overview

- A bit of history of economic thought
- How expectations are formed can matter *in the long run*
    - Seignorage model
- Learning without feedback
- Learning with feedback
    - Simple adaptive learning
    - Least-squares learning
    - Bayesian versus least-squares learning
    - Decision theoretic foundation of Adam & Marcet

Overview continued

Topics

- Learning & PEA
- Learning & sunspots

Why are expectations important?

- Most economic problems have intertemporal consequences
  - $\Longrightarrow$ future matters
- Moreover, future is uncertain
- Characteristics/behavior other agents can also be uncertain
  - $\Longrightarrow$ expectations can also matter in one-period problems

History of economic thought

- adaptive expectations:

$$\widehat{\mathsf{E}}_t\left[x_{t+1}\right] = \widehat{\mathsf{E}}_{t-1}\left[x_t\right] + \omega\left(x_t - \widehat{\mathsf{E}}_{t-1}\left[x_t\right]\right)$$

- very popular until the 70s

History of economic thought

problematic features of adaptive expectations:

- agents can be systematically wrong
- agents are completely passive:
  - $\widehat{\mathrm{E}}_t\left[x_{t+j}\right], j \geq 1$ only changes (at best) when x_t changes
  - $\Longrightarrow$ Pigou cycles are not possible
  - $\Longrightarrow$ model predictions *under*estimate speed of adjustment (e.g. for disinflation policies)

History of economic thought

problematic features of adaptive expectations:

- adaptive expectations about $x_{t+1} \neq$ adaptive expectations about Δx_{t+1}
  - (e.g. price level versus inflation)
- why wouldn't (some) agents use existing models to form expectations?
- expectations matter but still no role for randomness (of future realizations)
  - so no reason for buffer stock savings
  - no role for (model) uncertainty either

History of economic thought

rational expectations became popular because:

- agents are no longer passive machines, but forward looking
  - i.e., agents *think* through what could be consequences of their own actions and those of others (in particular government)
- consistency between model predictions and of agents being described
- randomness of future events become important
  - e.g., $\mathsf{E}_t\left[c_{t+1}^{-\gamma}\right] \neq \left(\mathsf{E}_t\left[c_{t+1}\right]\right)^{-\gamma}$

History of economic thought

problematic features of rational expectations

- agents have to know *complete* model
    - make correct predictions about all possible realizations
        - on *and* off the equilibrium path
- costs of forming expectations are ignored
- how agents get rational expectations is not explained

History of economic thought

problematic features of rational expectations

- makes analysis more complex
  - behavior this period depends on behavior tomorrow for all possible realizations
  - $\Longrightarrow$ we have to solve for policy *functions*, not just simulate the economy

Expectations matter

- Simple example to show that how expectations are formed can matter in the long run
  - See Adam, Evans, & Honkapohja (2006) for a more elaborate analysis

Model

- Overlapping generations
- Agents live for 2 periods
- Agents save by holding money
- No random shocks

Model

$$\max_{c_{1,t},c_{2,t}} \ln c_{1,t} + \ln c_{2,t}$$

s.t.

$$c_{2,t} \leq 1 + \frac{P_t}{P^e_{t+1}} (2 - c_{1,t})$$

no randomness $\implies$ we can work with expected value of variables instead of expected utility

Agent's behavior

First-order condition:

$$\frac{1}{c_{1,t}} = \frac{P_t}{P^e_{t+1}} \frac{1}{c_{2,t}} = \frac{1}{\pi^e_{t+1}} \frac{1}{c_{2,t}}$$

Solution for consumption:

$$c_{1,t} = 1 + \pi^e_{t+1}/2$$

Solution for real money balance (=savings):

$$m_t = 2 - c_{1,t} = 1 - \pi^e_{t+1}/2$$

Money supply

$$\overline{M}_t^s = \overline{M}$$

Equilibrium

Equilibrium in period t implies

$$
\begin{aligned}
\overline{M} &= M_t \\
\overline{M} &= P_t \left(1 - \pi^e_{t+1}/2\right) \\
P_t &= \frac{\overline{M}}{1 - \pi^e_{t+1}/2}
\end{aligned}
$$

Equilibrium

Combining with equilibrium in period $t-1$ gives

$$\pi_t = \frac{P_t}{P_{t-1}} = \frac{1-\pi_t^e/2}{1-\pi_{t+1}^e/2}$$

Thus: π_t^e & $\pi_{t+1}^e \Longrightarrow$ money demand $\Longrightarrow$ actual inflation π_t

Rational expectations solution

Optimizing behavior & equilibrium:

$$\frac{P_t}{P_{t-1}} = T(\pi_t^e, \pi_{t+1}^e)$$

Rational expectations equilibrium (REE):

$$\begin{aligned} \pi_t &= \pi_t^e \\ &\Longrightarrow \\ \pi_t &= T(\pi_t, \pi_{t+1}) \\ &\Longrightarrow \\ \pi_{t+1} &= 3 - \frac{2}{\pi_t} \\ \pi_{t+1} &= R\left(\pi_t\right) \end{aligned}$$

Multiple steady states

- There are two solutions to

$$\pi = 3 - \frac{2}{\pi}$$

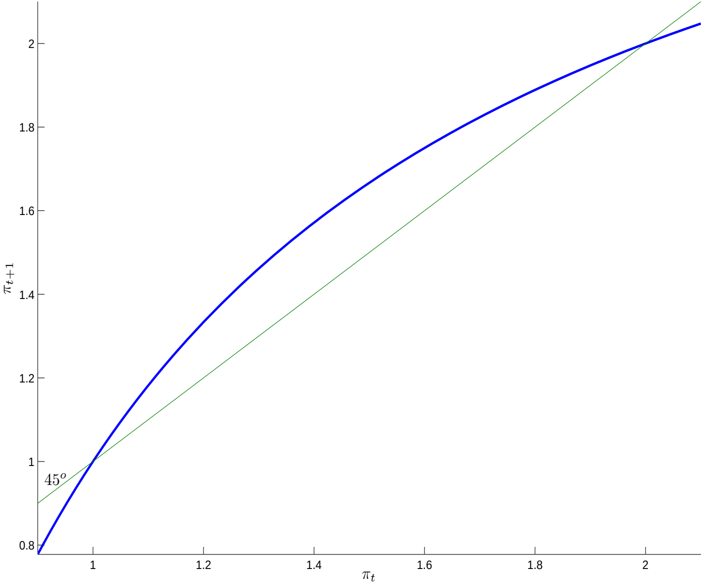
$\Longrightarrow$ there are two steady states

  - $\pi = 1$ (no inflation) and perfect consumption smoothing
  - $\pi = 2$ (high inflation), money has no value & no consumption smoothing at all

Unique solution

- Initial value for π_t not given, but given an initial condition the time path is fully determined
- π_t converging to 2 means m_t converging to zero and P_t converging to ∞

Rational expectations and stability



Rational expectations and stability

$$\begin{aligned}
\pi_1 &: \quad \text{value in period 1} \\
\pi_1 &< 1: \text{ divergence} \\
\pi_1 &= 1: \text{ economy stays at low-inflation steady state} \\
\pi_1 &> 1: \text{ convergence to high-inflation steady state}
\end{aligned}$$

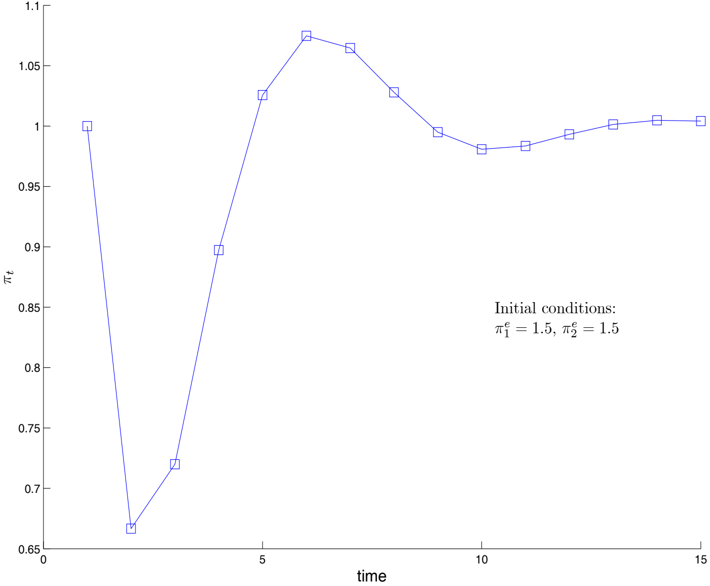
Alternative expectations

- Suppose that

$$\pi^e_{t+1} = \frac{1}{2}\pi_{t-1} + \frac{1}{2}\pi^e_t$$

- still the same two steady states, but
    - $\pi = 1$ is stable
    - $\pi = 2$ is not stable

Adaptive expectations and stability



Learning without feedback

Setup:

1. Agents know the complete model, except they do **not** know *dgp* exogenous processes
2. Agents use observations to update beliefs
3. Exogenous processes do not depend on beliefs $\Longrightarrow$ no feedback from learning to behavior of variable being forecasted

Learning without feedback & convergence

- If agents can learn the *dgp* of the exogenous processes, then you typically converge to REE
- They may not learn the correct *dgp* if
  - Agents use limited amount of data
  - Agents use misspecified time series process

Learning without feedback - Example

- Consider the following asset pricing model

$$P_t = \mathsf{E}_t \left[\beta \left(P_{t+1} + D_{t+1}\right)\right]$$

- If

$$\lim_{j \longrightarrow \infty} \beta^{t+j} D_{t+j} = 0$$

then

$$P_t = \mathsf{E}_t \left[\sum_{j=1}^{\infty} \beta^j D_{t+j}\right]$$

Learning without feedback - Example

- Suppose that

$$D_t = \rho D_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \sigma^2) \tag{1}$$

- REE:

$$P_t = \frac{D_t}{1 - \beta\rho}$$

(note that P_t could be negative so P_t is like a deviation from steady state level)

Learning without feedback - Example

- Suppose that agents do not know value of ρ
- Approach here:
  - If period t belief $= \widehat{\rho}_t$, then

$$P_t = \frac{D_t}{1 - \beta \widehat{\rho}_t}$$

  - Agents ignore that their beliefs may change,
    - i.e., $\widehat{\mathsf{E}}_t \left[P_{t+j} \right] = \mathsf{E}_t \left[\frac{D_{t+j}}{1-\beta \widehat{\rho}_{t+j}} \right]$ is assumed to equal $\frac{1}{1-\beta \widehat{\rho}_t} \mathsf{E}_t \left[D_{t+j} \right]$

Learning without feedback - Example

How to learn about ρ?

- Least squares learning using $\{D_t\}_{t=1}^{T}$ & correct *dgp*
- Least squares learning using $\{D_t\}_{t=1}^{T}$ & incorrect *dgp*
- Least squares learning using $\{D_t\}_{t=T-\bar{T}}^{T}$ & correct *dgp*
- Least squares learning using $\{D_t\}_{t=T-\bar{T}}^{T}$ & incorrect *dgp*
- Bayesian updating (also called rational learning)
- Lots of other possibilities

Convergence again

- Suppose that the true *dgp* is given by

$$
\begin{aligned}
D_t &= \rho_t D_{t-1} + \varepsilon_t \\
\rho_t &\in \left\{\rho_{\text{low}}, \rho_{\text{high}}\right\} \\
\rho_{t+1} &= \begin{cases} \rho_{\text{high}} \text{ w.p. } p(\rho_t) \\ \rho_{\text{low}} \text{ w.p. } 1 - p(\rho_t) \end{cases}
\end{aligned}
$$

- Suppose that agents think the true *dgp* is given by

$$D_t = \rho D_{t-1} + \varepsilon_t$$

- $\Longrightarrow$ Agents will never learn
  (see homework for importance of sample used to estimate ρ)

Recursive least-squares

- time-series model:

$$y_t = x_t'\gamma + u_t$$

- least-squares estimator

$$\widehat{\gamma}_T = R_T^{-1}\frac{X_T'Y_t}{T}$$

where

$$\begin{aligned} X_T' &= \begin{bmatrix} x_1 & x_2 & \cdots & x_T \end{bmatrix} \\ Y_T' &= \begin{bmatrix} y_1 & y_2 & \cdots & y_T \end{bmatrix} \\ R_T &= X_T'X_T/T \end{aligned}$$

Recursive least-squares

$$
\begin{aligned}
R_T &= R_{T-1} + \frac{\left(x_T x_T' - R_{T-1}\right)}{T} \\
\widehat{\gamma}_T &= \widehat{\gamma}_{T-1} + \frac{R_T^{-1} x_T \left(y_T - x_T' \widehat{\gamma}_{T-1}\right)}{T}
\end{aligned}
$$

Proof for R

$$\begin{aligned}
\frac{X_T'X_T}{T} &\stackrel{?}{=} \frac{X_{T-1}'X_{T-1}}{(T-1)} + \frac{x_Tx_T'}{T} - \frac{X_{T-1}'X_{T-1}}{T(T-1)} \\
\left(\frac{T-1}{T}\right) X_T'X_T &\stackrel{?}{=} X_{T-1}'X_{T-1} + \frac{T-1}{T}x_Tx_T' - \frac{X_{T-1}'X_{T-1}}{T} \\
X_T'X_T - \frac{X_T'X_T}{T} &\stackrel{?}{=} X_{T-1}'X_{T-1} + x_Tx_T' - \frac{x_Tx_T'}{T} - \frac{X_{T-1}'X_{T-1}}{T} \\
X_{T-1}'X_{T-1} + x_Tx_T' - \frac{X_{T-1}'X_{T-1} + x_Tx_T'}{T} &\stackrel{?}{=} X_{T-1}'X_{T-1} + x_Tx_T' - \frac{x_Tx_T' + X_{T-1}'X_{T-1}}{T}
\end{aligned}$$

Proof for gamma

$$
\begin{aligned}
\left(X_T' X_T\right)^{-1} \times X_T' Y_T &\overset{?}{=} \left(X_T' X_T\right)^{-1} \begin{pmatrix} \left(X_{T-1}' X_{T-1}\right)^{-1} X_{T-1}' Y_{T-1} + \\ x_T y_T \\ -x_T x_T' \left(X_{T-1}' X_{T-1}\right)^{-1} X_{T-1}' Y_{T-1} \end{pmatrix} \\
X_T' Y_T &\overset{?}{=} \left(X_{T-1}' X_{T-1} + x_T x_T'\right) \left(X_{T-1}' X_{T-1}\right)^{-1} X_{T-1}' Y_{T-1} \\
&\quad + \begin{pmatrix} x_T y_T \\ -x_T x_T' \left(X_{T-1}' X_{T-1}\right)^{-1} X_{T-1}' Y_{T-1} \end{pmatrix} \\
X_T' Y_T &\overset{?}{=} \left(I + x_T x_T' \left(X_{T-1}' X_{T-1}\right)^{-1}\right) X_{T-1}' Y_{T-1} \\
&\quad + \begin{pmatrix} x_T y_T \\ -x_T x_T' \left(X_{T-1}' X_{T-1}\right)^{-1} X_{T-1}' Y_{T-1} \end{pmatrix}
\end{aligned}
$$

Reasons to adopt recursive formulation

- makes proving analytical results easier
- less computer intensive,
  - but standard LS gives the same answer
- there are intuitive generalizations:

$$
\begin{aligned}
R_T &= R_{T-1} + \omega(T)\left(x_T x_T' - R_{T-1}\right) \\
\widehat{\gamma}_T &= \widehat{\gamma}_{T-1} + \omega(T) R_T^{-1} x_T \left(y_T - x_T' \widehat{\gamma}_{T-1}\right)
\end{aligned}
$$

$\omega(T)$ is the "gain"

Learning with feedback

1. Explanation of the idea
2. Simple adaptive learning
3. Least-squares learning
   - E-stability and convergence
4. Bayesian versus least-squares learning
5. Decision theoretic foundation of Adam & Marcet

Learning with feedback - basic setup

Model:

$$p_t = \rho \widehat{\mathsf{E}}_{t-1}\left[p_t\right] + \delta x_{t-1} + \varepsilon_t$$

RE solution:

$$\begin{aligned} p_t &= \frac{\delta}{1-\rho} x_{t-1} + \varepsilon_t \\ &= a_{\mathsf{re}} x_{t-1} + \varepsilon_t \end{aligned}$$

What is behind model

Model:

$$p_t = \rho \widehat{\mathsf{E}}_{t-1}\left[p_t\right] + \delta x_{t-1} + \varepsilon_t$$

Stories:

- Lucas aggregate supply model
- Muth market model

See Evans and Honkapohja (2009) for details

Learning with feedback - basic setup

Perceived law of motion (PLM) at $t-1$:

$$p_t = \widehat{a}_{t-1} x_{t-1} + \varepsilon_t \tag{2}$$

Actual law of motion (ALM):

$$p_t = \rho \widehat{a}_{t-1} x_{t-1} + \delta x_{t-1} + \varepsilon_t = \left(\rho \widehat{a}_{t-1} + \delta\right) x_{t-1} + \varepsilon_t \tag{3}$$

Updating beliefs I: Simple adaptive

$$\text{ALM:} \;\; p_t = (\rho \widehat{a}_{t-1} + \delta)\, x_{t-1} + \varepsilon_t$$

Simple adaptive learning:

- $\widehat{a}_t = \rho \widehat{a}_{t-1} + \delta$
- could be rationalized if
    - agents observe x_{t-1} and ε_t
    - t is more like an iteration and in each iteration agents get to observe long time-series to update

Simple adaptive learning: Convergence

$$\widehat{a}_t = \rho\widehat{a}_{t-1} + \delta$$

or in general

$$\widehat{a}_t - \widehat{a}_{t-1} = T\left(\widehat{a}_{t-1}\right)$$

Simple adaptive learning: Convergence

Key questions:

1. Does $\widehat{a}_t$ converge?
2. If yes, does it converge to a ?

Answers: If $|\rho| < 1$, then the answer to both is yes.

Updating beliefs: LS learning

Suppose agents use least-squares learning

$$
\begin{aligned}
\widehat{a}_t &= \widehat{a}_{t-1} + \frac{R_t^{-1} x_{t-1} \left(p_t - x_{t-1} \widehat{a}_{t-1} \right)}{t} \\
&= \widehat{a}_{t-1} + \frac{R_t^{-1} x_{t-1} \left(\left(\rho \widehat{a}_{t-1} + \delta \right) x_{t-1} + \varepsilon_t - x_{t-1} \widehat{a}_{t-1} \right)}{t} \\
R_t &= R_{t-1} + \frac{\left(x_{t-1} x_{t-1} - R_{t-1} \right)}{t}
\end{aligned}
$$

Updating beliefs: LS learning

$$
\begin{aligned}
\widehat{a}_t &= \widehat{a}_{t-1} + \frac{1}{t} R_t^{-1} x_{t-1} \left(p_t - x_{t-1} \widehat{a}_{t-1} \right) \\
&= \widehat{a}_{t-1} + \frac{1}{t} R_t^{-1} x_{t-1} \left(\left(\rho \widehat{a}_{t-1} + \delta \right) x_{t-1} + \varepsilon_t - x_{t-1} \widehat{a}_{t-1} \right) \\
R_t &= R_{t-1} + \frac{1}{t} \left(x_{t-1} x_{t-1} - R_{t-1} \right)
\end{aligned}
$$

To get system with only lags on RHS, let $R_t = S_{t-1}$

$$
\begin{aligned}
\widehat{a}_t &= \widehat{a}_{t-1} + \frac{1}{t} S_{t-1}^{-1} x_{t-1} \left(\left(\rho \widehat{a}_{t-1} + \delta \right) x_{t-1} + \varepsilon_t - x_{t-1} \widehat{a}_{t-1} \right) \\
S_t &= S_{t-1} + \frac{1}{t} \left(x_t x_t - S_{t-1} \right) \frac{t}{t+1}
\end{aligned}
$$

Updating beliefs: LS learning

Let

$$\theta_t = \begin{bmatrix} a_t \\ S_t \end{bmatrix}$$

Then the system can be written as

$$\begin{aligned} \widehat{\theta}_t &= \widehat{\theta}_{t-1} + \frac{1}{t} Q(\widehat{\theta}_{t-1}, x_t, x_{t-1}, \varepsilon_t) \\ & \text{or} \\ \Delta\widehat{\theta}_t &= T(\widehat{\theta}_{t-1}, x_t, x_{t-1}, \varepsilon_t, t) \end{aligned}$$

Note that

$$T(\cdot) = \frac{1}{t} Q(\cdot)$$

Key question

- If

$$\Delta\widehat{\theta}_t = \frac{1}{t} Q(\widehat{\theta}_{t-1}, x_t, x_{t-1}, \varepsilon_t, t)$$

then what can we "expect": about $\widehat{\theta}_t$?

- In particular, can we "expect" that

$$\lim_{t \longrightarrow \infty} \widehat{a}_t = a_{\mathsf{re}}$$

Corresponding differential equation

Much can be learned from following differential equation

$$\frac{d\theta}{d\tau} = h\left(\theta\left(\tau\right)\right)$$

where

$$h\left(\theta\right) = \lim_{t \to \infty} \mathsf{E}\left[Q(\theta, x_t, x_{t-1}, \varepsilon_t)\right]$$

Corresponding differential equation

In our example

$$
\begin{aligned}
h(\theta) &= \lim_{t\to\infty} \mathsf{E}\left[Q(\theta, x_t, x_{t-1}, \varepsilon_t)\right] \\
&= \lim_{t\to\infty} \mathsf{E}\left[\begin{array}{c} S^{-1}x_{t-1}\left(\left(\rho a+\delta\right)x_{t-1}+\varepsilon_t - x_{t-1}a\right) \\ \left(x_t x_t - S\right)\frac{t}{t+1} \end{array}\right] \\
&= \left[\begin{array}{c} MS^{-1}\left(\left(\rho-1\right)a+\delta\right) \\ M-S \end{array}\right]
\end{aligned}
$$

where

$$M = \lim_{t\to\infty} \mathsf{E}\left[x_t^2\right]$$

Analyze the differential equation

$$\frac{d\theta}{d\tau} = h\left(\theta\left(\tau\right)\right) = \left[\begin{array}{c} MS^{-1}\left(\left(\rho-1\right)a+\delta\right) \\ M-S \end{array}\right]$$

$$\frac{d\theta}{d\tau} = 0 \text{ if } M = S \;\&\; a = \frac{\delta}{1-\rho}$$

Thus, the (unique) rest point of $h\left(\theta\right)$ is the rational expectations solution

E-stability

$$\widehat{\theta}_t - \widehat{\theta}_{t-1} = \frac{1}{t} Q\left(\widehat{\theta}_{t-1}, x_t, x_{t-1}, \varepsilon_t, t\right)$$

Limiting behavior can be analyzed using

$$\frac{d\theta}{d\tau} = h(\theta(\tau)) = \lim_{t\to\infty} \mathsf{E}\left[Q(\theta, x_t, x_{t-1}, \varepsilon_t)\right]$$

A solution θ^*, e.g. $[a_{\mathsf{RE}}, M]$, is *"E-stable"* if $h(\theta)$ is stable at θ^*

E-stability

- $h(\theta)$ is stable if real part of the eigenvalues is negative:
- Here:

$$h(\theta) = \left[\begin{array}{c} (\rho - 1)\, a + \delta \\ M - S \end{array} \right]$$

$\Longrightarrow$ convergence of differentiable system if $\rho - 1 < 0$

  - $\Longrightarrow$ convergence even if $\rho < -1$!

Implications of E-stability?

- Recursive least-squares: stochastics in $T(\cdot)$ mapping
    - $\implies$ what will happen is less certain, even with E-stability

General implications of E-stability?

- If a solution **is not** E-stable:
  - $\Longrightarrow$ non-convergence is a probability 1 event
- If a solution **is** E-stable:
  - the presence of stochastics make the theorems non-trivial
  - in general only info about *mean dynamics*

Mean dynamics

See Evans and Honkapohja textbook for formal results.

- Theorems are a bit tricky, but are of the following kind:
  If a solution f^* is E-stable, then the time path under learning will either leave the neighborhood in finite time or will converge towards f^*. Moreover, the longer it does not leave this neighborhood, the smaller the probability that it will

- So there are two parts
  - mean dynamics: convergence towards fixed point
  - escape dynamics: (large) shocks may push you away from fixed point

Importance of Gain

$$\widehat{\gamma}_T = \widehat{\gamma}_{T-1} + \omega(T) R_T^{-1} x_T \left(y_T - x_T' \widehat{\gamma}_{T-1} \right)$$

- Gain in least squares updating formula, $\omega(T)$, plays a key role in theorems
- $\omega(T) \longrightarrow 0$ too fast: you may end up in somthing that is not an equilibrium
- $\omega(T) \longrightarrow 0$ too slowly:,you may not converge towards it
- So depending on the application, you may need conditions like

$$\sum_{t=1}^{\infty} \omega(t)^2 < \infty \text{ and } \sum_{t=1}^{\infty} \omega(t) = \infty$$

Special cases

- In simple cases, stronger results can be obtained
- Evans (1989) shows that the system of equations (2) and (3) with standard recursive least squares (gain of $1/t$) converges to rational expectations solution if $\rho < 1$ (so also if $\rho < -1$).

Bayesian learning

- LS learning has some disadvantages:
  - why "least-squares" and not something else?
  - how to choose gain?
  - why don't agents incorporate that beliefs change?
- Beliefs are updated each period
  $\Longrightarrow$ Bayesian learning is an obvious thing to consider

Bayesian versus LS learning

- LS learning $\neq$ Bayesian learning with uninformed prior
  at least not always
- Bullard and Suda (2009) provide following nice example

Bayesian versus LS learning

Model:

$$p_t = \rho_L p_{t-1} + \rho_0 \widehat{\mathsf{E}}_{t-1}[p_t] + \rho_1 \widehat{\mathsf{E}}_{t-1}[p_{t+1}] + \varepsilon_t \tag{4}$$

- Key difference with earlier model:
  - two extra terms

Bayesian versus LS learning

The RE solution:

- $$p_t = bp_{t-1} + \varepsilon_t$$

  where b is a solution to

  $$b = \rho_L + \rho_0 b + \rho_1 b^2$$

Bayesian learning - setup

- PLM:

$$p_t = \widehat{b}_{t-1} p_{t-1} + \varepsilon_t$$

  and ε_t has a known distribution

- plug PLM into (4) $\Longrightarrow$ ALM
  - but a Bayesian learner is a bit more careful

Bayesian learner understands he is learning

$$
\begin{aligned}
\widehat{\mathsf{E}}_{t-1}[p_{t+1}] &= \widehat{\mathsf{E}}_{t-1}\left[\rho_L p_t + \rho_0 \widehat{\mathsf{E}}_t[p_{t+1}] + \rho_1 \widehat{\mathsf{E}}_t[p_{t+2}]\right] \\
&= \rho_L p_t + \widehat{\mathsf{E}}_{t-1}\left[\rho_0 \widehat{\mathsf{E}}_t[p_{t+1}] + \rho_1 \widehat{\mathsf{E}}_t[p_{t+2}]\right] \\
&= \rho_L p_t + \widehat{\mathsf{E}}_{t-1}\left[\rho_0 \widehat{b}_t p_t + \rho_1 \widehat{b}_t p_{t+1}\right]
\end{aligned}
$$

- and he realizes, for example, that $\widehat{b}_t$ and p_t are both affected by ε_t!

Bayesian learner understands he is learning

- Bayesian learner realizes that

$$\widehat{\mathsf{E}}_{t-1}\left[\widehat{b}_t p_{t+1}\right] \neq \widehat{\mathsf{E}}_{t-1}\left[\widehat{b}_t\right]\widehat{\mathsf{E}}_{t-1}\left[p_{t+1}\right]$$

  and calculates $\widehat{\mathsf{E}}_{t-1}\left[\widehat{b}_t p_{t+1}\right]$ explicitly

- LS learner forms expectations thinking that

$$\begin{aligned}\widehat{\mathsf{E}}_{t-1}\left[\widehat{b}_t p_{t+1}\right] &= \widehat{\mathsf{E}}_{t-1}\left[\widehat{b}_{t-1} p_{t+1}\right] \\ &= \widehat{b}_{t-1}\widehat{\mathsf{E}}_{t-1}\left[\left(\rho_L p + \rho_0 \widehat{b}_{t-1} + \rho_1 \widehat{b}_{t-1}\right) p_t\right]\end{aligned}$$

Bayesian versus LS learning

- Bayesian learner cares about a covariance term
- Bullard and Suda (2009) show that Bayesian is simillar to LS learning in terms of E-stability
- Such covariance terms more important in nonlinear frameworks
- Unfortunately not much done with nonlinear models

Learning what?

Model:

$$P_t = \beta \mathsf{E}_t \left[P_{t+1} + D_{t+1} \right]$$

- Learning can be incorporated in many ways.
- Obvious choices here:
    1. learn about *dgp* D_t and use true mapping for $P_t = P(D_t)$
    2. know *dgp* D_t and learn about $P_t = P(D_t)$
    3. learn about both

Learning what?

1. Adam, Marcet, Nicolini (2009): one can solve several asset pricing puzzles using a simple model if learning is learning about $\mathsf{E}_t\left[P_{t+1}\right]$ (instead of learning about *dgp* D_t)

2. Adam and Marcet (2011): provide micro foundations that this is a sensible choice

Simple model

Model:

$$P_t = \beta \mathsf{E}_t \left[P_{t+1} + D_{t+1} \right]$$

$$\frac{D_{t+1}}{D_t} = a\varepsilon_{t+1}$$

with

$$\mathsf{E}_t \left[\varepsilon_{t+1} \right] = 1$$

ε_t i.i.d.

Model properties REE

- Solution:

$$P_t = \frac{\beta a}{1 - \beta a} D_t$$

- P_t / D_t is constant
- P_t / P_{t-1} is i.i.d.

Adam, Marcet, & Nicolini 2009

PLM:

$$\widehat{\mathsf{E}}_t\left[\frac{P_{t+1}}{P_t}\right] = \gamma_t$$

ALM:

$$\begin{aligned} \frac{P_t}{P_{t-1}} &= \frac{1-\beta\gamma_{t-1}}{1-\beta\gamma_t} a\varepsilon_t = \left(a + \frac{a\beta\Delta\gamma_t}{1-\beta\gamma_t}\right)\varepsilon_t \\ \gamma_{t+1} &= \left(a + \frac{a\beta\Delta\gamma_t}{1-\beta\gamma_t}\right) \end{aligned}$$

Model properties with learning

- Solution is quite nonlinear
  - especially if γ_t is close to β^{-1}
- Serial correlation.
  - in fact there is momentum. For example:

$$\begin{aligned}\gamma_t &= a \;\&\; \Delta\gamma_t > 0 \Longrightarrow \Delta\gamma_{t+1} > 0\\ \gamma_t &= a \;\&\; \Delta\gamma_t < 0 \Longrightarrow \Delta\gamma_{t+1} < 0\end{aligned}$$

- P_t/D_t is time varying

Adam, Marcet, & Nicolini 2011

Agent i does following optimization problem

$$\max \widehat{\mathsf{E}}_{i,t}\left[\cdot\right]$$

- $\widehat{\mathsf{E}}_{i,t}$ is based on a sensible probability measure
- $\widehat{\mathsf{E}}_{i,t}$ is not necessarily the true conditional expectation

Adam, Marcet, & Nicolini 2011

- Setup leads to standard first-order conditions but with $\widehat{\mathsf{E}}_{i,t}$ instead of E_t
- For example

$$P_t = \beta \widehat{\mathsf{E}}_{i,t}\left[P_{t+1} + D_{t+1}\right]$$

if agent i is not constrained

- Key idea:
  - price determination is difficult
  - agents do not know this mapping
  - $\Longrightarrow$ they forecast $\widehat{\mathsf{E}}_{i,t}\left[P_{t+1}\right]$ directly
  - $\Longrightarrow$ law of iterated expectations cannot be used because next period agent i may be constrained in which case the equality does not hold

Topics - Overview

1. E-stability and sunspots
2. Learning and nonlinearities
   Parameterized expectations
3. Two representations of sunspots

E-stability and sunspots

Model:

$$x_t = \rho \mathsf{E}_t \left[x_{t+1} \right]$$

x_t cannot explode

no initial condition

Solution:

$$\begin{aligned} |\rho| &< 1 : x_t = 0 \ \forall t \\ |\rho| &\geq 1 : x_t = \rho^{-1} x_{t-1} + e_t \ \forall t \end{aligned}$$

where e_t is the sunspot (which has $\mathsf{E}_t \left[e_t \right] = 0$

Adaptive learning

PLM:

$$x_t = \widehat{a}_t x_{t-1} + e_t$$

ALM:

$$\begin{aligned} x_t &= \widehat{a}_t \rho x_{t-1} \\ &\Longrightarrow \widehat{a}_{t+1} = \widehat{a}_t \rho \end{aligned}$$

- thus divergence when $|\rho| > 1$ (sunspot solutions)

Adaptive learning

PLM:

$$x_t = \widehat{a}_t x_{t-1} + e_t$$

ALM:

$$\begin{aligned} x_t &= \widehat{a}_t \rho x_{t-1} \\ &\Longrightarrow \widehat{a}_{t+1} = \widehat{a}_t \rho \end{aligned}$$

- thus divergence when $|\rho| > 1$ (sunspot solutions)

Stability puzzle

- There are few counter examples and not too clear why sunspots are not learnable in RBC-type models
  - sunspot solutions are learnable in some New Keynesian models (Evans and McGough 2005)
  - McGough, Meng, and Xue 2011 provide a counterexample and show that an RBC model with *negative* externalities has learnable sunspot solutions

PEA and learning

- Learning is usually done in linear frameworks
- PEA parameterized the conditional expectations in nonlinear frameworks
- $\Longrightarrow$ PEA is a natural setting to do
  - adaptive learning as well as
  - recursive learning

Model

$$P_t \quad = \quad \mathsf{E}\left[\beta\left(\frac{D_{t+1}}{D_t}\right)^{-\nu}(P_{t+1}+D_{t+1})\right] = G\left(X_t\right)$$

X_t : state variables

Conventional PEA in a nutshell

- Start with a guess for $G\left(X_t\right)$, say $g(x_t;\eta_0)$
  - $g\left(\cdot\right)$ may have wrong functional form
  - x_t may only be a subset of X_t
  - η_0 are the coefficients of $g\left(\cdot\right)$

Conventional PEA in a nutshell

- Iterate to find fixed point for η_i

1. use η_i to generate time path $\{P_t\}_{t=1}^{T}$
2. let

$$\hat{\eta}_i = \arg\min_{\eta} \sum_t \left(y_{t+1} - g\left(x_t; \eta\right)\right)^2$$

where

$$y_{t+1} = \beta \left(\frac{D_{t+1}}{D_t}\right)^{-\nu} \left(P_{t+1} + D_{t+1}\right)$$

3. Dampen if necessary

$$\eta_{i+1} = \omega \hat{\eta}_i + (1 - \omega)\, \eta_i$$

Interpretation of conventional PEA

- Agents have beliefs
- Agents get to observe long sample generated with these beliefs
- Agents update beliefs
- Corresponds to adaptive expectations
    - no stochastics if T is large enough

Recursive PEA

- Agents form expectations using $g\left(x_t; \eta_t\right)$

- Solve for P_t using

$$P_t = g\left(x_t; \eta_t\right)$$

- Update beliefs using this one additional observation

- Go to the next period using η_{t+1}

Recursive methods and convergence

Look at recursive formulation of LS:

$$\widehat{\gamma}_t = \widehat{\gamma}_{t-1} + \frac{1}{t} R_t^{-1} x_t \left(y_t - x_t' \widehat{\gamma}_{t-1} \right)$$

- !!! $\Delta\hat{\gamma}_t$ gets smaller as t gets bigger

General form versus common factor represenation

sunspot literature distinquishes between:

1. General form representation of a sunspot
2. Common factor representation of a sunspot

First consider non-sun-spot indeterminacy

Model:

$$\begin{aligned} k_{t+1} + a_1 k_t + a_2 k_{t-1} &= 0 \text{ or} \\ (1 - \lambda_1 L)(1 - \lambda_2 L) k_{t+1} &= 0 \end{aligned}$$

Also:

- k_0 given
- k_t has to remain finite

Multiplicity

Solution:

$$\begin{aligned} k_t &= b_1\lambda_1^t + b_2\lambda_2^t \\ k_0 &= b_1 + b_2 \end{aligned}$$

Thus many possible choices for b_1 and b_2 if $|\lambda_1| < 1$ and $|\lambda_1| < 2$

Multiplicity

- What if we impose recursivity?

$$k_t = \bar{d}k_{t-1}$$

- Does that get rid of multiplicity? No, but it does reduce the number of solutions from ∞ to 2

$$\begin{aligned}
\left(\bar{d}^2 + a_1\bar{d} + a_2\right) k_{t-1} &= 0 \ \ \forall t \\
&\Longrightarrow \\
\left(\bar{d}^2 + a_1\bar{d} + a_2\right) &= 0
\end{aligned}$$

the 2 solutions correspond to setting either λ_1 or λ_2 equal to 0

Back to sunspots

Doing the same trick with sunspots gives a solution with following two properties:

1. it has a *serially correlated* sunspot component
   with the same factor as the endogenous variable (i.e. the common factor)
2. there are two of these

General form representation

Model:

$$\begin{aligned} \mathsf{E}_t\left[k_{t+1} + a_1 k_t + a_2 k_{t-1}\right] &= 0 \quad \text{or} \\ \mathsf{E}_t\left[\left(1-\lambda_1 L\right)\left(1-\lambda_2 L\right) k_{t+1}\right] &= 0 \end{aligned}$$

General form representation:

$$\begin{aligned} k_t &= b_1 \lambda_1^t + b_2 \lambda_2^t + e_t \\ k_0 &= b_1 + b_2 + e_0 \end{aligned}$$

where e_t is serially uncorrelated

Common factor representation

Model:

$$
\begin{aligned}
\mathsf{E}_t\left[k_{t+1} + a_1 k_t + a_2 k_{t-1}\right] &= 0 \quad \text{or} \\
\mathsf{E}_t\left[(1-\lambda_1 L)(1-\lambda_2 L)\, k_{t+1}\right] &= 0
\end{aligned}
$$

Common factor representation:

$$
\begin{aligned}
k_t &= b_1 \lambda_i^t + \zeta_t \\
\zeta_t &= \lambda_i \zeta_{t-1} + e_t \\
k_0 &= b_i + \zeta_0 \\
\lambda_i &\in \{\lambda_1, \lambda_2\}
\end{aligned}
$$

where e_t is serially uncorrelated

References

- Adam, K., G. Evans, and S. Honkapohja, 2006, Are hyperinflation paths learnable, Journal of Economic Dynamics and Control.
    - Paper shows that the high-inflation steady state is stable under learning in the seignorage model as discussed in slides.
- Adam, K., A. Marcet, and J.P. Nicolini, 2009, Stock market volatility and learning, manuscript.
    - Paper shows that learning about endogenous variables like prices gives you much more "action" than learning about exogenous processes (i.e. they show that learning with feedback is more interesting than learning without feedback).

References

- Adam, K., and A. Marcet, 2011, Internal rationality, imperfect market knowledge and asset prices, Journal of Economic Theory.
  - Paper motivates that the thing to be learned is the conditional expectation in the Euler equation.
- Bullard, J. and J. Suda, 2009, The stability of macroeconomics systems with Bayesian learners, manuscript.
  - Paper gives a nice example on the difference between Bayesian and Least-Squares learning.

References

- Evans, G.W., and S. Honkapohja, 2001, Learning and expectations in macroeconomics.
  - Textbook on learning dealing with all the technical stuff very carefully.
- Evans, G.W., and S. Honkaphja, 2009, Learning and macroeconomics, Annual Review of Economics.
  - Survey paper.
- Evans, G.W., and B. McGough, 2005, Indeterminacy and the stability puzzle in non-convex economies, Contributions to macroeconomics.
  - Paper argues that sunspot solutions (or indeterminate solutions more generally) are not stable (not learnable) in RBC-type models.
- Evans, G.W., and B. McGough, 2005, Monetary policy, indeterminacy, and learning, Journal of Economic Dynamics and Control.
  - Paper shows that the NK model with some Taylor rules has learnable sunspot solutions
- McGough, B., Q. Meng, and J. Xue, 2011, Indeterminacy and E-stability in real business cycle models with factor-generated externalities, manuscript.
  - Paper provides a nice example of a sunspot in a linearized RBC-type model that