Learning to Optimize

George W. Evans

University of Oregon and University of St Andrews

Bruce McGough

University of Oregon

Workshop on Adaptive Learning

May 7-8, 2018,

Bilbao - Spain

Outline

– Introduction

– Shadow-price learning for dynamic programming problems

– Learning to optimize in an LQ problem

– Value function learning & Euler-equation learning

– Applications: R Crusoe model; Market Equilibrium Investment model

– Conclusions

Introduction

- In microfounded models we assume agents are rational in two ways:
  – they form forecasts optimally (they are endowed with RE)
  – they make choices by maximizing their objective function

- RE may be implausibly demanding. The adaptive (e.g. least-squares) learning approach is a natural bounded-rationality response to this critique. See, e.g., Marcet & Sargent (1989), Evans & Honkapohja (2001).

- Under least-squares learning, agents can learn over time to coordinate on an REE in "self-referential models" if it is "E-stable." Interesting learning dynamics can emerge.

- That agents are endowed with the solution to dynamic optimization problems is equally implausible: it may take time to learn to optimize.

- Boundedly optimal decision-making is a natural complement to boundedly rational forecasting. It obeys the "cognitive consistency principle."

- Our implementation, which we call shadow-price learning, complements and extends least-squares learning in expectation formation.

- Using shadow-price learning agents can learn over time to solve their dynamic stochastic optimization problem.

- Again, interesting learning dynamics can emerge.

Literature on agent-level learning and decision-making

– Cogley and Sargent (IER, 2008). Bayesian decision-making with learning.

– Adam and Marcet (JET, 2011). "Internal rationality."

– Preston (IJCB, 2005). Eusepi & Preston (AEJmacro 2010), 'Anticipated utility' and infinite-horizon decisions. Evans, Honkapohja & Mitra (JME, 2009).

– Evans and Honkapohja (ScandJE, 2006) and Honkapohja, Mitra and Evans (2013). "Euler-equation learning." Howitt and Özak (2014).

– Watkins (1989). Q-learning. 'Quality value' of state-action pairs. Typical applications are to models with finite states and actions.

– Marimon, McGrattan and Sargent (JEDC, 1990). Classifier systems. Lettau and Uhlig (AER, 1999).

Shadow-price learning

We now introduce our approach – Shadow-price (SP) learning. Consider a standard dynamic programming problem

$$\begin{aligned} V^*(x_0) &= \max E_0 \sum_{t \geq 0} \beta^t r(x_t, u_t) \\ \text{subject to } x_{t+1} &= g(x_t, u_t, \varepsilon_{t+1}) \end{aligned}$$

and $\bar{x}_0$ given, with $u_t \in \Gamma(x_t) \subseteq \mathbb{R}^m$ and $x_t \in \mathbb{R}^n$.

Linear-Quadratic (LQ) special case:

$$\begin{aligned} r(x_t, u_t) &= x_t' R x_t + u_t' Q u_t + 2 x_t' W u_t \\ g(x_t, u_t, \varepsilon_{t+1}) &= A x_t + B u_t + C \varepsilon_{t+1}. \end{aligned}$$

Examples:

1. Robinson Crusoe problem:

$$\max -E \sum_{t \geq 0} \beta^t \left((c_t - b^*)^2 + \phi s_t^2 \right),$$

where $b^* > 0$ and $\phi > 0$, subject to

$$\begin{aligned} y_t &= A_1 s_t + A_2 s_{t-1} \\ s_{t+1} &= y_t - c_t + \mu_{t+1} \end{aligned}$$

Output y_t is fruit/sprouting trees. c_t = consumption. Trees live two years. s_t = number of young trees at t = number of old trees at $t+1$. Young trees need weeding.

2. Investment under uncertainty: Lucas-Prescott-Sargent model of investment. Firms sell goods in a competitive market. Market demand is

$$p_t = \alpha_0 - \alpha_1 y_t + v_t,$$

where v_t is AR(1) stationary and observable. Firm ω's problem is

$$\begin{aligned}
\max_{I_t(\omega)} \quad & \hat{E}(\omega) \sum_{t \geq 0} \beta^t \left(p_t y_t(\omega) - (J + q_t(\omega)) I_t(\omega) - \frac{\gamma}{2} I_t(\omega)^2 \right) \\
k_t(\omega) &= (1-\delta) k_{t-1}(\omega) + \mu I_t(\omega) + (1-\mu) I_{t-1}(\omega), \text{ with } 0 < \delta \leq 1 \\
y_t(\omega) &= k_t(\omega)^\alpha \text{ where } 0 < \alpha \leq 1, \\
p_t &= a_0 + a_1 p_{t-1} + a_2 v_{t-1} + \varepsilon_t^p,
\end{aligned}$$

where $q_t(\omega) = q(\omega) + \varepsilon_t^q(\omega)$. Firm's problem is LQ for $\alpha = 1$.

Three cases: (i) Single agent LQ problem for given price process. (ii) Single agent non-LQ problem. (ii) Market equilibrium with p_t given by demand and $y_t = \int_\Omega y_t(\omega) d\omega$.

Shadow-price learning

Return to consideration of the standard dynamic programming problem

$$\begin{aligned} V^*(x_0) &= \max E_0 \sum_{t\geq 0} \beta^t r(x_t, u_t) \\ \text{subject to } x_{t+1} &= g(x_t, u_t, \varepsilon_{t+1}) \end{aligned}$$

and $\bar{x}_0$ given, with $u_t \in \Gamma(x_t) \subseteq \mathbb{R}^m$ and $x_t \in \mathbb{R}^n$. The state x_t includes an intercept. Our approach is based on the corresponding Lagrangian

$$\mathcal{L} = E_0 \sum_{t\geq 0} \beta^t \left(r(x_t, u_t) + \lambda_t^{*\prime} \left(g(x_{t-1}, u_{t-1}, \varepsilon_t) - x_t \right) \right).$$

Our starting point is the FOC and envelope condition

$$\begin{aligned} 0 &= r_u(x_t, u_t)' + \beta E_t g_u(x_t, u_t, \varepsilon_{t+1})' \lambda_{t+1}^*. \\ \lambda_t^* &= r_x(x_t, u_t)' + \beta E_t g_x(x_t, u_t, \varepsilon_{t+1})' \lambda_{t+1}^* \end{aligned}$$

In **SP learning** we replace λ_t^* with λ_t, the perceived shadow price of the state x_t, and we treat these equations as **behavioral**.

To implement this we need forecasts. In line with the adaptive learning literature $x_{t+1} = g(x_t, u_t, \varepsilon_{t+1})$ is often assumed **unknown** and is **approximated** by

$$x_{t+1} = Ax_t + Bu_t + C\varepsilon_{t+1},$$

where unknown parameter estimates are updated over time using RLS, i.e. recursive LS. Agents must also forecast λ_{t+1}. We assume that they believe the dependence of λ_t on x_t can be approximated by

$$\lambda_t = Hx_t + \mu_t,$$

where estimates of H are updated over time using RLS.

To implement SP learning, given x_t and estimates A, B, H the agent sets u_t to satisfy

$$r_u(x_t, u_t)' = -\beta B' \hat{E}_t \lambda_{t+1}, \text{ since } B = \partial x_{t+1}/\partial u_t, \text{ and}$$

$$\hat{E}_t \lambda_{t+1} = H\left(Ax_t + Bu_t\right)$$

for u_t and $\hat{E}_t \lambda_{t+1}$. The FOC r_u equation may in general be nonlinear.

The x_t FOC (envelope condition) gives a value for λ_t

$$\lambda_t = r_x(x_t, u_t)' + \beta A' \hat{E}_t \lambda_{t+1}, \text{ since } A = \partial x_{t+1}/\partial x_t,$$

which is used next period to update the estimate of H. This equation has an asset price interpretation.

At $t+1$ RLS is used to update estimates of A and/or B and H in

$$x_{t+1} = Ax_t + Bu_t + C\varepsilon_{t+1} \text{ and } \lambda_t = Hx_t + \mu_t,$$

This fully defines SP learning as a recursive system.

Advantages of SP learning as a model of boundedly optimal decision-making:

- The **pivotal role of shadow prices** λ_t, central to economic decisions.
- $\hat{E}_t \lambda_{t+1}$ and transition dynamics $B = \partial x_{t+1}/\partial u_t$ measure the **intertemporal trade-off** which determines actions u_t.
- **Simplicity**. Agents each period solve a **two-period problem** – an attractive level of sophistication.
- Incorporates **recursive LS updating** of A, B, H, the hallmark of **adaptive learning**, but extended to include forecasts of shadow prices.

- As we will see, although our agents are **boundedly optimal**, in a LQ setting they **become fully optimal asymptotically**.

- SP learning **can be incorporated into** standard **DSGE models**.

We also outline two alternative implementations of SP learning:

– Value function learning: value function estimated instead of shadow prices.

– Euler equation learning: closely related to SP-learning in special cases.

SP learning is related to the other approaches in the literature:

– Like Q-learning and classifier systems, it builds off of Bellman's equation.

– Like Internal Rationality we do not impose RE.

– As with anticipated utility/IH agents neglect parameter uncertainty.

– Like Euler-equation learning, it is sufficient to forecast one step ahead.

– Like anticipated utility/IH & Internal Rationality, an agent-level approach

SP-learning has simplicity, generality and economic intuition, and can be embedded in general equilibrium models with heterogeneous agents..

Simplified Robinson Crusoe example

- Illustrate SP-learning using Crusoe with $y_t = A_1 s_t,$, i.e. $A_2 = 0$. Then

$$\max -E \sum_{t \geq 0} \beta^t \left((c_t - b^*)^2 + \phi s_t^2 \right), \text{ s.t. } s_{t+1} = A_1 s_t - c_t + \mu_{t+1}$$

- Since $r_c = 2\left(b^* - c_t\right)$ the **control decision** for c_t satisfies

$$2\left(b^* - c_t\right) = \beta \hat{E}_t \lambda_{t+1}$$

for $t+1$ SP of tree λ_{t+1}. PLM $\lambda_t = H_0 + H_1 s_t \rightarrow$

$$\hat{E}_t \lambda_{t+1} = H_{0,t-1} + H_{1,t-1}\left(A_{1,t-1} s_t - c_t\right)$$

Given s_t **these two equations can be solved for** c_t and $\hat{E}_t \lambda_{t+1}$.

- Next step: how to update estimates of A_1 and $H = (H_0, H_1)$?

- To update $A_{1,t-1}$ to A_{1t} we add data point (s_t, s_{t-1}) and use LS to update estimate of $s_t = A_1 s_{t-1} - c_{t-1} + \mu_t$.

- To update H_{t-1} to H_t we first use $r_s = -2\phi s_t$ to **compute estimate**

$$\lambda_t = -2\phi s_t + \beta A_{1,t-1} \hat{E}_t \lambda_{t+1}.$$

  Then we add data point (λ_t, s_t) and use LS to update the estimate of $\lambda_t = H_0 + H_1 s_t + \eta_t$.

- This is now a fully specified recursive system.

Learning to Optimize in an LQ set-up

- To prove results we specialize the dynamic programming set-up to be the standard linear-quadratic set-up, which has been extensively studied and widely applied. In this set-up we can obtain our asymptotic convergence result.

- Consider the single-agent problem: determine a sequence of controls u_t that solve, given the initial state x_0,

$$\begin{aligned} V^*(x_0) = \max \quad & -E_0 \sum \beta^t \left(x_t' R x_t + u_t' Q u_t + 2 x_t' W u_t \right) \\ s.t. \quad & x_{t+1} = A x_t + B u_t + C \varepsilon_{t+1}. \end{aligned}$$

  We make standard assumptions on R, Q, W, A, B: LQ.1 (concavity), LQ.2 (stabilizability) and LQ.3 (detectability).

- Under LQ1 – LQ3 the optimal controls are given by

$$u_t = -F^* x_t \text{ where } F^* = \left(Q + \beta B' P^* B\right)^{-1} (\beta B' P^* A + W')$$

where P^* is obtained by analyzing Bellman's equation and satisfies

$$P^* = R + \beta A' P^* A - (\beta A' P^* B + W) \left(Q + \beta B' P^* B\right)^{-1} (\beta B' P^* A + W').$$

Also $V^*(x) = -x' P^* x - \beta(1-\beta)^{-1} \mathsf{tr}(\sigma_\varepsilon^2 P^* CC')$.

- Solving the "Riccati equation" for P^* generally only possible numerically. This requires a sophisticated agent with a lot of knowledge and computational skills. Our agents follow a simpler boundedly optimal procedure.

- Our approach replaces RE and full optimality with adaptive learning and bounded optimality, based on shadow prices.

- For LQ models the true transition equation is linear and the optimal shadow price equation is linear.

- The SP-learning system can be written recursively as:

$$
\begin{aligned}
x_t &= Ax_{t-1} + Bu_{t-1} + C\varepsilon_t \\
\mathcal{R}_t &= \mathcal{R}_{t-1} + \gamma_t \left(x_t x_t' - \mathcal{R}_{t-1} \right) \\
H_t' &= H_{t-1}' + \gamma_t \mathcal{R}_{t-1}^{-1} x_{t-1} \left(\lambda_{t-1} - H_{t-1} x_{t-1} \right)' \\
A_t' &= A_{t-1}' + \gamma_t \mathcal{R}_{t-1}^{-1} x_{t-1} \left(x_t - Bu_{t-1} - A_{t-1} x_{t-1} \right)' \\
u_t &= F^{SP}(H_t, A_t, B) x_t \\
\lambda_t &= T^{SP}(H_t, A_t, B) x_t \\
\gamma_t &= t^{-1} \text{ or } \gamma_t = \kappa(t+N)^{-1}
\end{aligned}
$$

  In this formulation A is estimated but B is assumed known, which would be typical. WE use RLS, the recursive formulation of LS.

- For real-time learning results we need an additional assumption, LQ.RTL: the state dynamics are well-behaved under optimal decision-making, i.e. are stationary and have a non-singular second-moment matrix.

Theorem 4 (Asymptotic optimality of SP learning in LQ model). *If LQ.1 - LQ.3 and LQ.RTL are satisfied then, locally,* (H_t, A_t) *converges to* (H^*, A) *almost surely when the recursive algorithm is augmented with a suitable projection facility, and* $F^{SP}(H_t, A_t, B)$ *converges to* $-F^*$.

Extension: We show it is unnecessary for agents to estimate and forecast shadow prices for *exogenous* states. This is convenient for applications.

Proof of Theorem 4

The proof of Theorem 4 is given in the paper. The ingredients of the proof are:

- We use well-known results for recursive stochastic algorithms (SRAs) that are widely used in the adaptive learning literature.

- These results show that under suitable assumptions, which we show are satisfied, (H_t, A_t) locally converges almost surely to (H^*, A) provided an associated differential differential equation (ODE) is locally stable.

- We then use known properties of LQ problems to show that assumptions LQ1-LQ3 imply that the ODE is indeed locally stable.

Theorem 4 is a striking result:

- SP learning converges asymptotically to fully rational forecasts and fully optimal decisions.

- By including perceived shadow prices, we have converted an infinite-horizon problem into a two-period optimization problem.

- The agent is learning over its lifetime based on a single 'realization' of its decisions and the resulting states.

Remark 1: The "projection facility" in many applications is rarely needed.

Remark 2: like adaptive learning of expectations, the system is self-referential. Here this comes from the impact of perceived shadow prices on actual decisions.

Alternative implementation: value-function learning

- In SP learning agents estimate the SP vector λ for state x. An alternative implementation is to estimate $V(x)$ to make decisions using

$$V(x) = -x' \hat{P}_t x.$$

- They use $\hat{P}_t$ and the rhs of Bellman's equation to obtain revised $\hat{V}_t(x_t)$

$$\hat{V}_t(x_t) = -\left(x_t' R x_t + u_t' Q u_t + 2 x_t' W u_t\right) + \beta \hat{E}_t x_{t+1} \hat{P}_t x_{t+1}.$$

- Estimates $\hat{P}_t$ of P are updated over time using a regression of $\hat{V}_t$ on linear and quadratic terms in the state x_t.

- Theorem 5 provides a corresponding result for value-function learning.

Alternative implementation: Euler-equation learning

- Another alternative implementation of bounded optimality is EE learning.

- One-step-ahead Euler eqns exist in special cases. If x_{t+1} does not depend on endogenous states x_t,then λ_t can be eliminated to give

$$Qu_t + W'x_t + \beta B'E_t(Rx_{t+1} + Wu_{t+1}) = 0.$$

- Under EE-learning agents use this to make decisions using a forecast of its own future decision u_{t+1} based on $u_t = -Fx_t$, i.e.

$$u_{t+1} = -F_t\hat{E}_t x_{t+1}.$$

  Theorem 6 provides a corresponding result for EE learning.

Example: SP Learning in a Crusoe economy

$$\max -E \sum_{t \geq 0} \beta^t \left((c_t - b^*)^2 + \phi s_{t-1}^2 \right)$$

$$\text{s.t.} \quad s_{t+1} = A_1 s_t + A_2 s_{t-1} - c_t + \mu_{t+1}$$

Output is fruit/sprouting trees. Young trees need weeding. Under SP learning Bob estimates the SPs of young and old trees:

$$\lambda_{it} = a_i + b_i s_t + d_i s_{t-1}, \text{ for } i = 1, 2, \text{ and thus}$$

$$\hat{E}_t \lambda_{it+1} = a_i + b_i (A_{1t} s_t + A_{2t} s_{t-1} - c_t) + d_i s_t, \text{ for } i = 1, 2.$$

These plus the FOC for the control

$$c_t = b^* - \frac{\beta}{2} \hat{E}_t \lambda_{1t+1}.$$

determine $c_t, E_t \lambda_{1,t+1}, E_t \lambda_{2,t+1}$, given s_t, s_{t-1}.

The FOCs for the states give updated estimates of SPs

$$\begin{aligned}\lambda_{1t} &= -2\phi s_t + \beta A_{1t}\hat{E}_t\lambda_{1t+1} + \beta\hat{E}_t\lambda_{2t+1} \\ \lambda_{2t} &= \beta A_{2t}\hat{E}_t\lambda_{1t+1},\end{aligned}$$

which allows Bob to use RLS update the SP equation coefficients.

Proposition: *Provided LQ.RTL holds, Robinson Crusoe learns to optimally consume fruit.*

Note: LQ.RTL necessarily holds if $A_2 \geq 0$ is not too large and shocks have small support.

EE learning is also possible using a second-order Euler equation. See paper.

Example: SP Learning in the Investment Model

Recall the firm ω problem

$$\begin{aligned} \max_{I_t(\omega)} \quad & \hat{E}(\omega) \sum_{t \geq 0} \beta^t \left(p_t y_t(\omega) - (J + q_t(\omega)) I_t(\omega) - \frac{\gamma}{2} I_t(\omega)^2 \right) \\ k_t(\omega) \;\; = \;\; & (1-\delta) k_{t-1}(\omega) + \mu I_t(\omega) + (1-\mu) I_{t-1}(\omega), \text{ with } 0 < \delta \leq 1 \end{aligned}$$

where $y_t(\omega) = k_t(\omega)^{\alpha}$. The firm treats the price process p_t as exogenous. It is convenient to rewrite the problem in terms of *installed capital*

$$z_t(\omega) = (1-\delta) k_{t-1}(\omega) + (1-\mu) I_{t-1}(\omega),$$

where $k_t(\omega) = z_t(\omega) + \mu I_t(\omega)$.

The firm problem is then

$$\max_{I_t(\omega)} \hat{E}(\omega) \sum_{t \geq 0} \beta^t \left(p_t f(z_t(\omega) + \mu I_t(\omega)) - (J + q_t(\omega)) I_t(\omega) - \frac{\gamma}{2} I_t(\omega)^2 \right)$$

$$\begin{aligned} z_{t+1}(\omega) &= (1-\delta) z_t(\omega) + (1 - \delta\mu) I_t(\omega) \\ v_{t+1} &= \rho_v v_t + \varepsilon^v_{t+1}, \; q_{t+1} = \bar{q} + \varepsilon^q_{t+1} \\ p_{t+1} &= a_0 + a_1 p_t + a_2 v_t + \varepsilon^p_{t+1} \end{aligned}$$

Exogenous and endogenous states for the firm are: $x_{1t} = (1, v_t, q_t(\omega), p_t)$, $x_{2t} = z_t(\omega)$, and the control takes the form

$$u_t = I_t(\omega) = I\left(p_t, v_t, q_t\left(\omega\right), z_t(\omega)\right).$$

Below we give the details for SP-learning in this model, in which case $I\left(\cdot\right)$ at t depends on H_t, a_t, i.e. estimates of SP and state transition parameters.

Market temporary equilibrium

The adaptive learning literature uses a **temporary equilibrium** approach (Hicks, 1939): given expectations and perceptions (SPs), however formed, firms make time-t decisions, conditional on prices.

The price p_t is then determined by market clearing. Recalling market demand

$$p_t = \alpha_0 - \alpha_1 y_t + v_t,$$

p_t, investment $I_t(\omega)$, and installed capacity $z_{t+1}(\omega)$ are determined by

$$\begin{aligned} p_t &= \alpha_0 - \alpha_1 \int_\Omega f\left(z_t(\omega) + \mu I(p_t, v_t, q_t(\omega), z_t(\omega))\right) d\omega + v_t \\ z_{t+1}(\omega) &= (1-\delta) z_t(\omega) + (1-\delta\mu) I(p_t, v_t, q_t(\omega), z_t(\omega)). \end{aligned}$$

Rational expectations

For the LQ case $\alpha = 1$ the market equilibrium price process takes the form

$$p_{t+1} = a_0 + a_1 p_t + a_2 v_t + \varepsilon^p_{t+1}$$

and the solution, which is linear, can be obtained. For $\alpha \in (0, 1)$ the economy is nonlinear but we can approximate the REE to first order.

The IRFs for the demand shock are given in Fig. 3. For $\alpha = 0.5$, diminishing returns results in a smaller initial supply response, leading to a larger initial increase in p, which subsequently induces greater I. As p returns to its steady state level, I temporarily falls below its steady-state level.

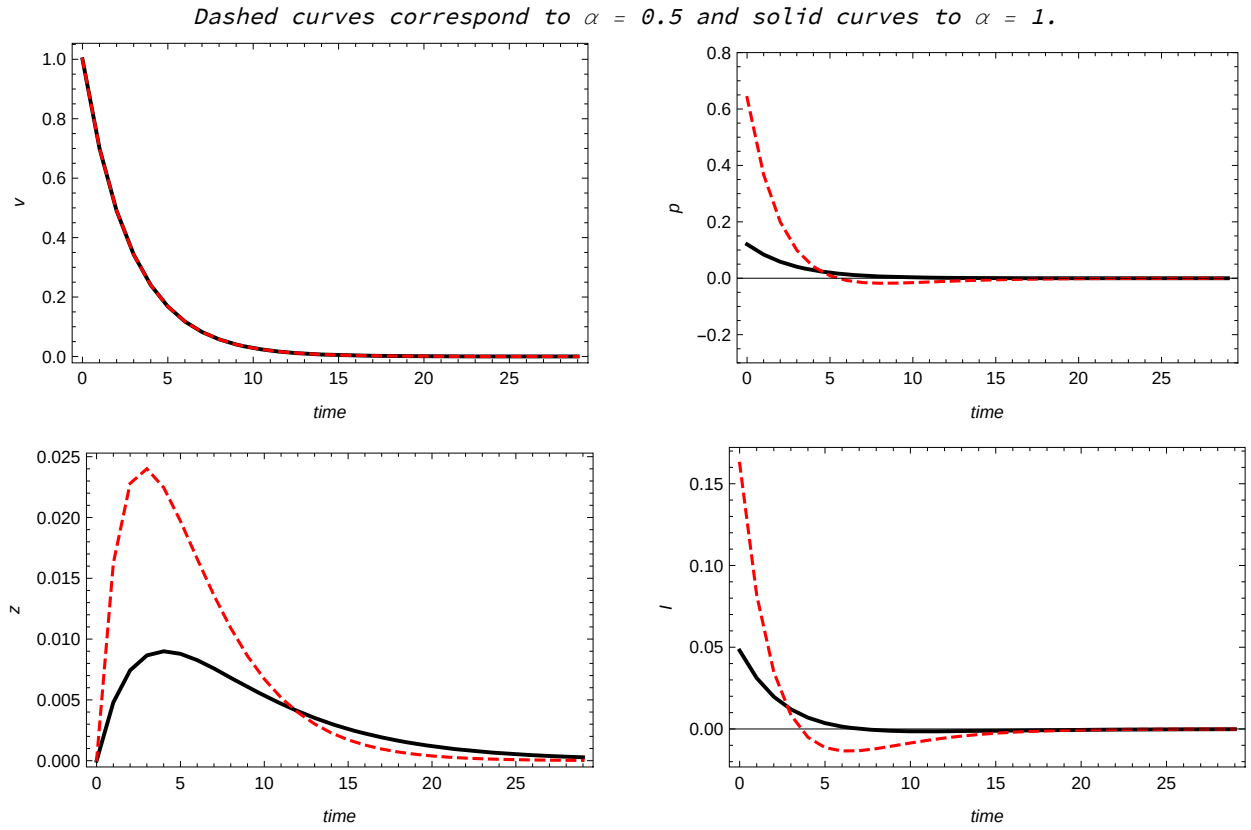


Figure 3: IRFs for demand shock in the REE

Shadow–price learning: firm problem

Under SP-learning firm ω has estimate $\lambda_t^z(\omega)$ for the SP at t of an additional unit of installed capital. Omitting ω, the FOC for I_t gives behavioral equation

$$J + q_t(\omega) + \gamma I_t(\omega) = \mu p_t f'(z_t(\omega) + \mu I_t(\omega)) + \beta(1-\delta\mu)\hat{E}_t\lambda_{t+1}^z(\omega).$$

Decision-making requires also $\hat{E}_t\lambda_{t+1}^z$, using forecasts of the state at $t+1$, including $\hat{E}_t p_{t+1}$ based on estimates of a_0, a_1, a_2, and estimates of

$$\lambda_t^z = H_0 + H_1 v_t + H_2 q_t + H_3 p_t + H_4 z_t.$$

The FOC for z_t

$$\lambda_t^z = p_t f'(z_t + \mu I_t) + \beta(1-\delta)\hat{E}_t\lambda_{t+1}^z$$

gives a new data point for λ_t^z for updating estimates H.

To illustrate, we give numerical results for the LQ case $f(k) = k^{\alpha}$ with $\alpha = 1$, and for a non-LQ case with $\alpha = 0.3$, where p_t is exogenous and consistent with the REE, and with the price process parameters assumed known. The learning gain parameter is 0.1 and the other parameters were set at

$$\alpha_0 = 10, \alpha_1 = -1.1, \beta = .95, \ \mu = .95, \ \delta = .1, \ J = 2, \ \bar{q} = 0, \ \gamma = 2, \rho = .7.$$

Figures 4 and 5 show real-time plots of estimated SP parameters H. These strongly suggest convergence (to first-order in the non-LQ case) to optimal decision-making: agents learn how to optimize using SP-learning.

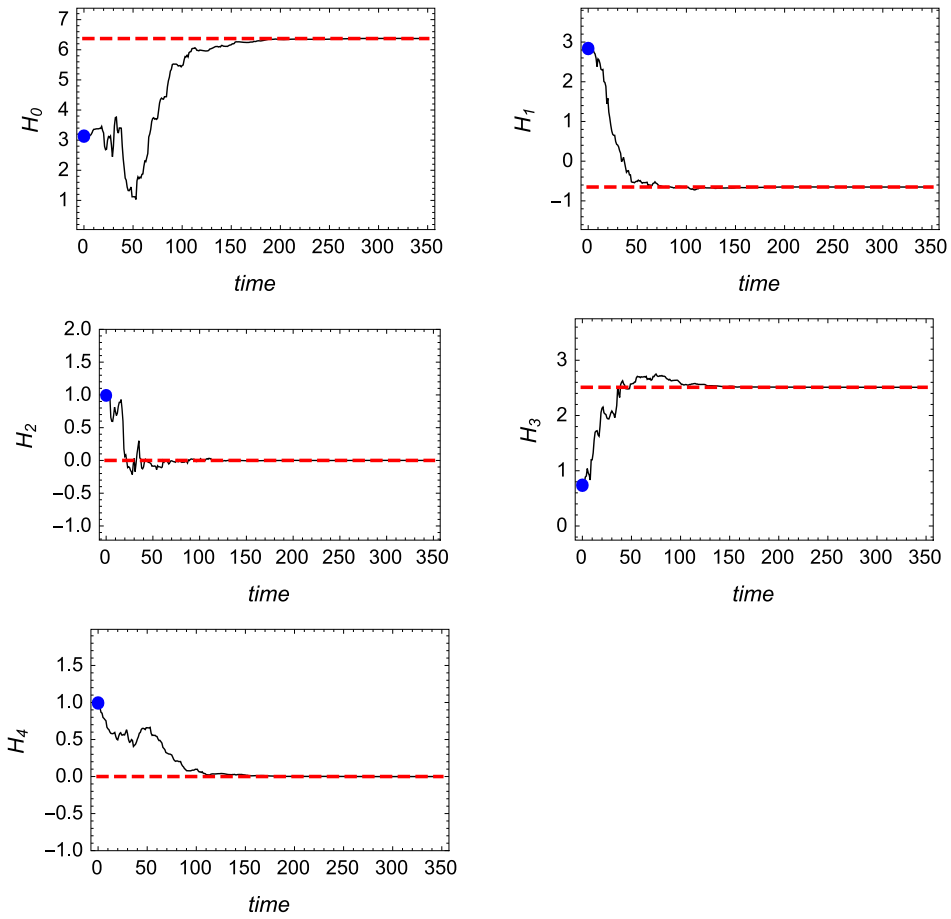


Fig. 4: Beliefs H for SP with exogenous goods price. LQ case.

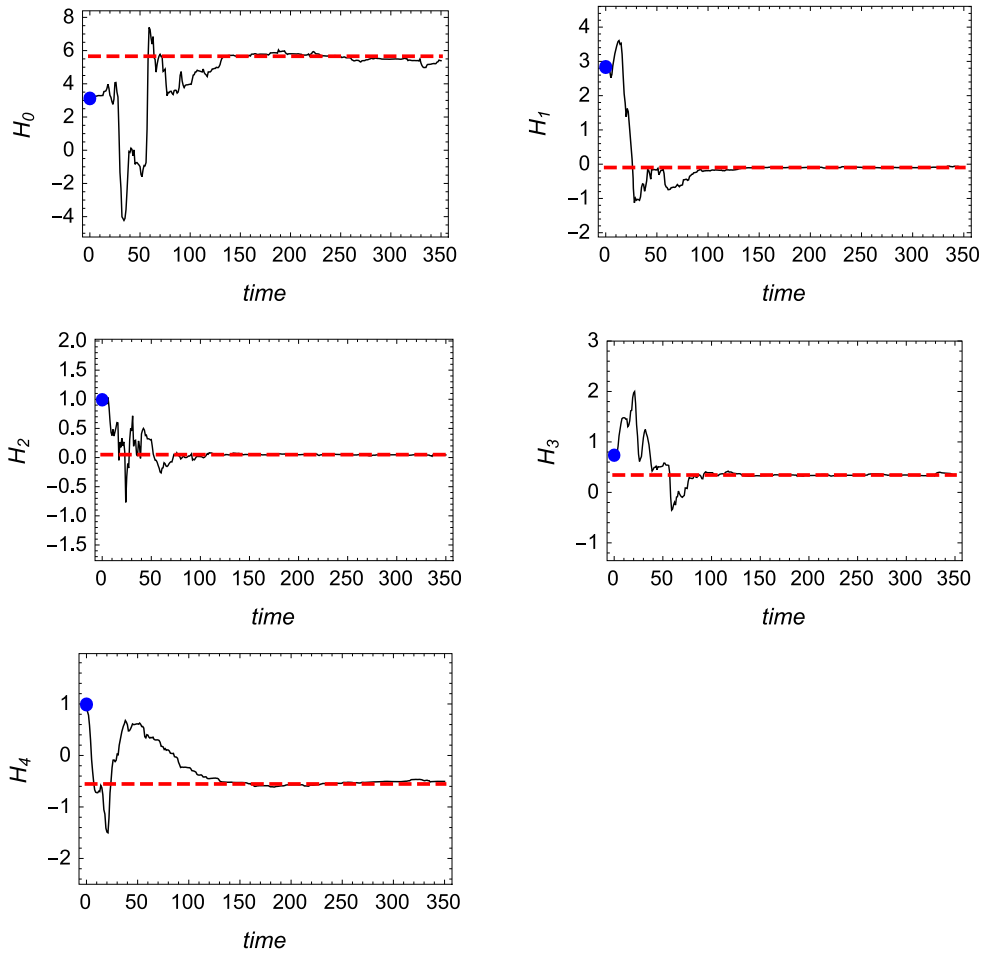


Figure 5: Beliefs H for SP with exogenous price process. Non-LQ case.

SP-learning in Market Equilibrium

- Now we embed SP-learners in a market equilibrium setting.

- Can firms over time learn both to forecast correctly and to make optimal dynamic decisions?

- We restrict attention to the LQ case $\alpha = 1$ and the representative agents case and for simplicity assume also $q_t = 0$.

- With homogeneous agents, in the REE the components of the agent's state vector $x_t = (1, v_t, p_t, z_t)$ is collinear, so to forecast the shadow price we assume agents use the subvector $\tilde{x}_t = (1, p_t, z_t)$.

- Let H_t be time t estimate of regressing $\{\lambda_s\}$ on $\{\tilde{x}_s = (1, p_s, z_s)\}$ and a_t be the t estimate from regressing $\{p_s\}$ on $\{1, p_{s-1}, v_{s-1}\}$.

- (H_{t-1}, a_{t-1}) summarizes the agent's beliefs for decision-making at t.

Writing $\hat{x}_t = (1, v_t, z_t)$ and $x_t = (1, v_t, p_t, z_t)$, we have conditional decisions

$$\begin{aligned} I_t &= I(x_t, H_{t-1}, A_{t-1}) = I(\hat{x}_t, p_t, H_{t-1}, a_{t-1}) \\ \lambda_t &= \lambda(x_t, H_{t-1}, A_{t-1}) = \lambda(\hat{x}_t, p_t, H_{t-1}, a_{t-1}). \end{aligned}$$

The $\mathcal{TE}$ map is defined implicitly as the price that clears the goods market:

$$p_t = \alpha_0 - \alpha_1 f(z_t + I(\hat{x}_t, p_t, H_{t-1}, a_{t-1})) + v_t \implies p_t = \mathcal{TE}(\hat{x}_t, H_{t-1}, a_{t-1}).$$

Fig. 6 gives results of a typical simulation (dashed red shows REE values). There is apparent convergence to REE and optimal decision-making.

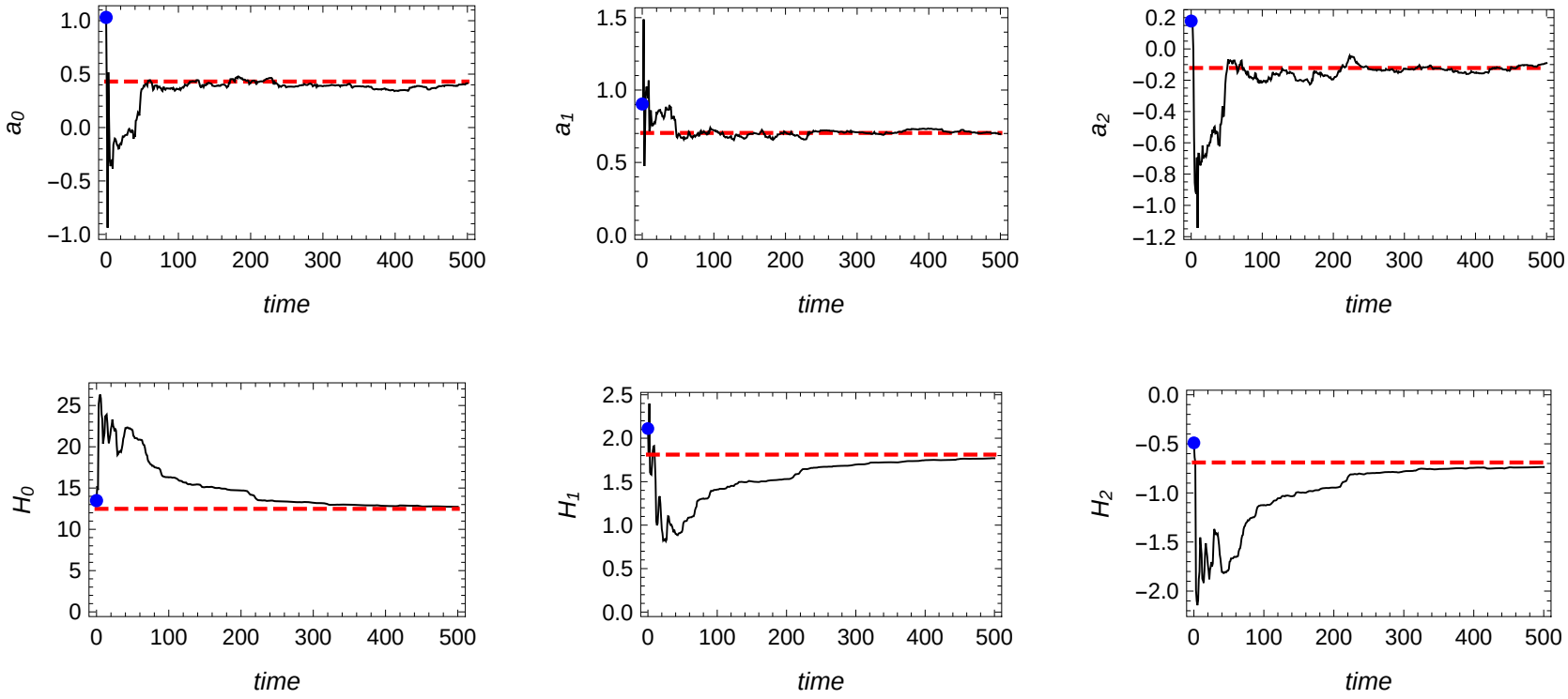


Figure 6: Beliefs parameters in market equilibrium. LQ case $\alpha = 1$. Top panel: market-price parameters. Bottom panel: shadow-price parameters.

Conclusions

- SP learning can be applied in general dynamic stochastic optimization problems and within general equilibrium models.

- The approach is formulated at the agent level and allows for heterogeneity in general equilibrium settings.

- It is tractable because agents need only solve 2-period optimization problems using one-step ahead forecasts of states and shadow prices.

- SP learning is boundedly optimal but converges to optimal decisions.

- Current work – Applications:

  – SP learning in DSGE models with heterogeneous agents.

  – Develop general procedure for implementing SP-learning in such settings.

  Extensions:

  – SP learning with inequality constraints (e.g. borrowing constraints).

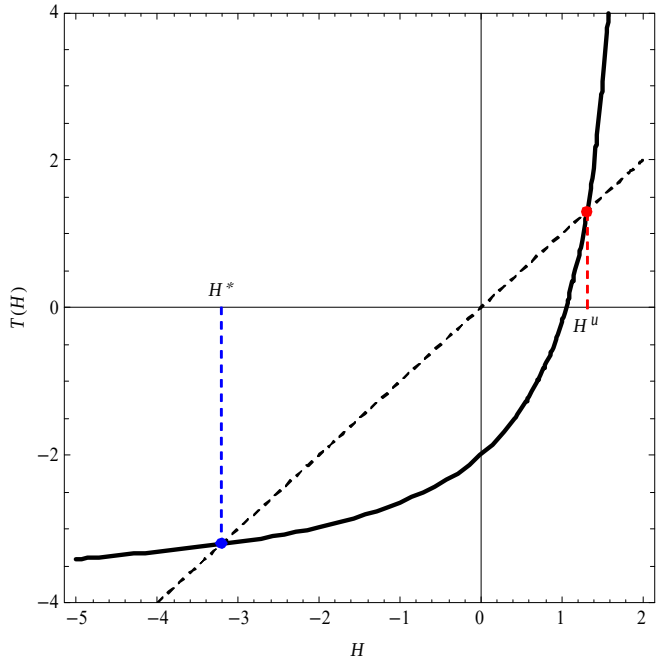
  – Value function learning in qualitative choice models.

  Study implications of persistent deviations from full optimization due to:

  1. misspecified shadow price models
  2. persistent learning dynamics, "escape paths" under constant gains.
  3. Inequality arising from heterogenous SP priors.

ADDENDA – Projection Facility

Illustration of need for Projection Facility (PF): Without PF an unusual sequence of shocks can lead to perceptions H that impart explosive dynamics.



Univariate T-map.

Numerical results confirm instability arises only when stability conditions are only barely satisfied, e.g. in a simple 2-dimensional model, instability without PF arises frequently if

$$\rho\left(DT_H\left(H^*, A, B\right)\right) = .9916 \text{ and } \rho\left(A + BF\left(H^*, A, B\right)\right) = .9966$$

where $\rho\left(\cdot\right)$ is the spectral radius, whereas if

$$\rho\left(DT_H\left(H^*, A, B\right)\right) = .8387 \text{ and } \rho\left(A + BF\left(H^*, A, B\right)\right) = .8429$$

even with constant gain $\gamma_t = 0.01$, instability never arose in 500 simulations over 2500 periods and all 500 converged toward and stayed near the optimum.

LQ assumptions

The paper shows the model can be transformed to

$$\begin{aligned} \max \quad & -\sum \left(\hat{x}_t' \hat{R} \hat{x}_t + \hat{u}_t' Q \hat{u}_t \right) \\ s.t. \quad & \hat{x}_{t+1} = \hat{A} \hat{x}_t + \hat{B} \hat{u}_t, \end{aligned}$$

where

$$\begin{aligned} \hat{R} &= R - W Q^{-1} W' \\ \hat{A} &= \beta^{\frac{1}{2}} \left(A - B Q^{-1} W' \right) \\ \hat{B} &= \beta^{\frac{1}{2}} B. \end{aligned}$$

LQ.1: (Concavity). The matrix $\hat{R}$ is symmetric positive semi-definite and the matrix Q is symmetric positive definite.

$\hat{R}$ can be factored as $\hat{R} = \hat{D}\hat{D}'$, where rank$(\hat{R}) = r$ and $\hat{D}$ is $n \times r$. With this notation, we say that:

- A matrix is stable if its eigenvalues have modulus less than one.

- The matrix pair $(\hat{A}, \hat{B})$ is *stabilizable* if there exists a matrix K such that $\hat{A} + \hat{B}K$ is stable.

- The matrix pair $(\hat{A}, \hat{D})$ is *detectable* provided that whenever y is a (nonzero) eigenvector of $\hat{A}$ associated with the eigenvalue μ and $\hat{D}'y = 0$ it follows that $|\mu| < 1$. Intuitively, $\hat{D}'$ acts as a factor of the objective function's quadratic form $\hat{R}$: if $\hat{D}'y = 0$ then y is not detected by the objective function; in this case, the associated eigenvalue must be contracting.

With these definitions in hand, we may formally state the assumptions we make concerning the matrices identifying the LQ problem.

LQ.2: The system $(\hat{A}, \hat{B})$ is stabilizable.

LQ.3: The system $(\hat{A}, \hat{D})$ is detectable.

Intuitively, **concavity** ensures the instantaneous objective function is bounded from above. **Stabilizability** ensures that there is a bounded control sequence such that the trajectory of the state is also bounded, i.e. avoiding unbounded paths is feasible. **Detectability** ensues that avoiding unbounded paths is desirable for the agent (because it implies that, whenever the state gets large in magnitude, the instantaneous objective gets large in magnitude.

Assumption LQ.RTL

For asymptotic convergence of real-time learning dynamics we need the additional assumption:

LQ.RTL The eigenvalues of $A + BF(H^*, A, B)$ not corresponding to the constant term have modulus less than one, and the associated asymptotic second-moment matrix for the process $x_t = (A + BF(H^*, A, B))\, x_{t-1} + C\varepsilon_t$ is non-singular.

Intuitively, LQ.RTL states that the state dynamics are well-behaved under optimal decision-making, i.e. are stationary and have a non-singular second-moment matrix: the state dynamics are non-explosive and do not exhibit asymptotic perfect multicollinearity.

Euler-equation learning in Crusoe model

EE learning is possible using a second-order Euler equation:

$$\begin{aligned} c_t - \beta\phi\hat{E}_t s_{t+1} &= \Psi_t + \beta A_{1t}\hat{E}_t c_{t+1} + \beta^2 A_{2t}\hat{E}_t c_{t+2}, \\ \text{where } \Psi_t &= b^*(1 - \beta A_{1t} - \beta^2 A_{2t}). \end{aligned}$$

To implement use forecasts of $\hat{E}_t c_{t+i}$ from estimates of

$$c_t = a_3 + b_3 s_t + d_3 s_{t-1}.$$

SP learning and EE-learning are not identical, but both are asymptotically optimal. This can be seen from a numerical calculation of their largest eigenvalue, shown in Figure 2.

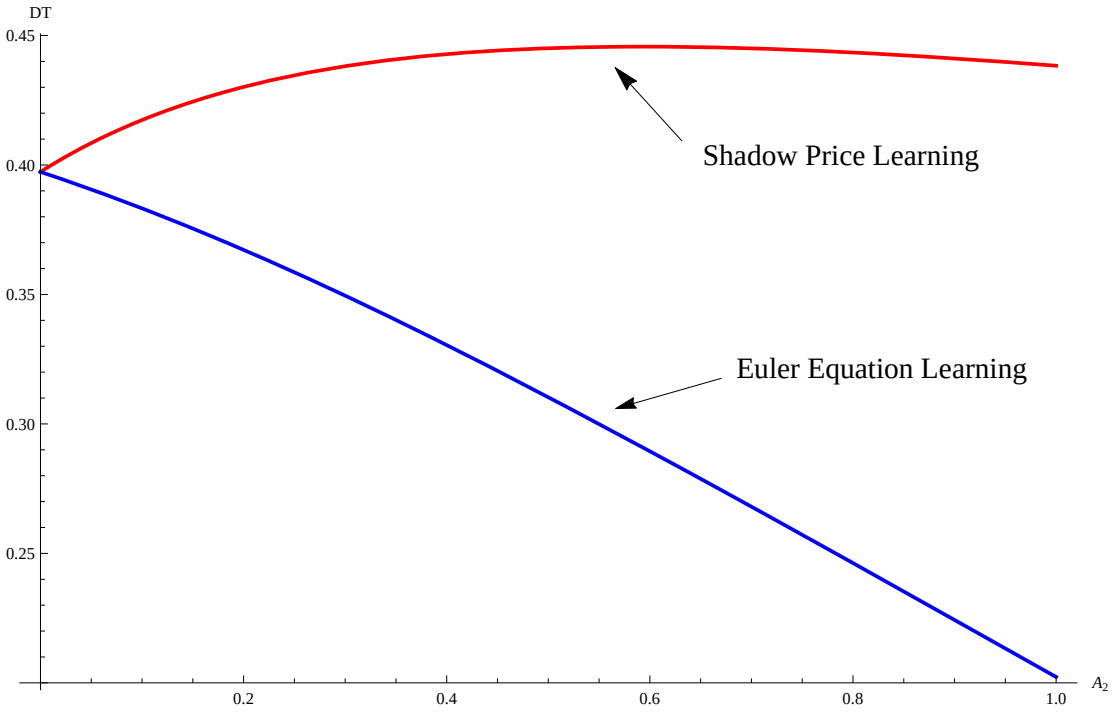


Figure 2: Largest eigenvalue of DT^{SP} and DT^{EL} under SP and EE learning.

Why are EE-learning and SP learning different?

Here $\dim(u) = 1$ and $\dim(x) = 2$. The PLMs are

$$\text{SP PLM: } \lambda_t = Hx_t \text{ vs EE PLM: } u_t = Fx_t$$

so SP learning estimates 4 parameters whereas EE learning estimates 2 parameters.

The SP PLM requires less structural information than the EE PLM. For the SP PLM to be equivalent to the EE PLM, agents would need to understand the structural relation between λ_1 and λ_2 and to impose this restriction in estimation.