Learning to Optimize

George W. Evans

Uinversity of Oregon and University of St Andrews

Bruce McGough

University of Oregon

CDMA Workshop on Macroeconomic Policy and Expectations

University of St Andrews,

12 September, 2013

Outline

- Introduction
- Literature
- Shadow-price learning
- Learning to optimize in an LQ problem
- Euler-equation learning
- Application to Crusoe model
- Application to Ramsey 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

- The critique of Rational Expectations (RE) – that it is often implausibly demanding – is well-known. The adaptive (e.g. least-squares) learning approach is a natural bounded-rationality response to this critique.

- In our view the assumption that agents are endowed with the solution to their dynamic optimization problem is equally implausible: just as it may take time for agents to learn to form RE, so it may take time for them to learn to optimize.

- Thus boundedly optimal decision-making is a natural complement to boundedly rational forecasting.

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

- Our central result is to show that by using shadow-pice learning, agents can learn over time to solve their dynamic stochastic optimization problem.

Literature on agent-level learning and decision-making

- Cogley and Sargent (IER, 2008). Bayesian decision-making in a permanent-income model. Even in a finite planning model with 2-state Markov income process, the optimal decision rule requires considerable sophistication.

- Adam and Marcet (JET, 2011). "Internal rationality." Agents solve their dynamic optimization problem for a given system of probability beliefs on future paths of variables exogenous to their decisions. For risk-neutral application agents make decisions based on one-step-ahead forecasts.

- Preston (IJCB, 2005). Using the 'anticipated utility' approach, agents fully solve their optimal decision-making problem at t, given their time t forecast at of the whole futiure path of variables exogenous to them, but ignoring that their estimated model will change over time. Examples include Eusepi and Preston (AEJmacro 2010) and Evans, Honkapohja and Mitra (JME, 2009).

- Evans and Honkapohja (ScandJE, 2006) and Honkapohja, Mitra and Evans (2013). "Euler-equation learning." Agents make decisions based on their Euler equation, using one-step ahead forecasts of the relevant varibles including their own next-period decisions, e.g., rates of return and their own future consumption. See also Howit and Özak (2009).

- Watkins (1989). Q-learning. Based on the Bellman's equation, agents estimate and update the 'quality value' of state-action pairs. Typical applications are to models with finite states and actions.

- Marimon, McGrattan and Sargent (JEDC, 1990) use a related approach based on classifier systems. Lettau and Uhlig (AER, 1999) incorporate rules of thumb into dynamic programming using classifier systems.

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 $x_t \in \mathbb{R}^n$. Our approach is based on the corresponding Lagrangian

$$\mathcal{L} = E_0 \sum_{t \geq 0} \beta^t \left(r(x_t, u_t) + \lambda_t' \left(g(x_{t-1}, u_{t-1}, \varepsilon_t) - x_t \right) \right).$$

Our starting point is the FOCs

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

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 assumed unknown and is approximated by

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

where estimates of A, B, C are updated over time using 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.

SP-learning is thus specified by solving simultaneously the u_t FOC and the $\hat{E}_t\lambda^*_{t+1}$ forecast equation

$$r_u(x_t, u_t)' = -\beta B' \hat{E}_t\lambda^*_{t+1} \text{ and } \hat{E}_t\lambda^*_{t+1} = H\left(Ax_t + Bu_t\right)$$

for u_t and $\hat{E}_t\lambda^*_{t+1}$. These can then be used with the x_t FOC for to obtain an updated estimate of λ^*_t

$$\hat{E}_t\lambda^*_t = r_x(x_t, u_t)' + \beta A' \hat{E}_t\lambda^*_{t+1}.$$

At $t+1$ RLS is used to update the estimates of A, B, H in

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + C\varepsilon_{t+1}, \\ \hat{E}_t\lambda^*_t &= Hx_t + \mu_t, \end{aligned}$$

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^*, which are central to economic decisions.

- $\hat{E}_t \lambda_{t+1}^*$ and transition dynamics $B = \partial x_{t+1} / \partial u_t'$ measures the intertemporal trade-off which determines actions u_t.

- Simplicity. Agents act as if they solve a two-period problem - an attractive level of sophistication.

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

- Incorporates RLS updating of A, B, H, the hallmark of adaptive learning, but extended to include forecasts of shadow prices.

- Can incorporate this model of bounded optimality into standard DSGE models.

SP-learning is related to the alternative approaches:

– Euler equation learning can be viewed as a special case when $\textsf{dim}(u_t) = \textsf{dim}(x_t)$.

– Like Q-learning and classifier systems, it builds off of the intuition of Bellman's equation. But instead of trying to learn $V(x)$ our agents try to learn $\lambda(x) = V'(x)$.

– Like infinite-horizon learning we use the anticipated utility approach, not the more sophisticated Bayesian approach

– Although the alternative approaches have some advantages, we find SP-learning attractive due to its simplicity, generality and economic intuition, while at the same time delivering asymptotic convergence to fully optimal decision-making.

Learning to Optimize in an LQ set-up

- We now specialize the dynamic programming set-up to be the standard linear-quadratic set-up, which has been extensively used 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,

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

  A simple example is a linear-quadratic Robinson Crusoe economy.

- Under well-known conditions the sequence of controls are determined by

$$u_t = Fx_t \text{ where } F = -\left(Q + \beta B'PB\right)^{-1}(\beta B'PA + W')$$

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

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

We state this well-known result as Theorem 1.

- Solving this "Riccati equation" is generally only possible numerically. This requires a sophisticated agent with a lot of knowledge and computational skills. Under our approach agents follow a simpler boundedly optimal procedure.

- In our approach we replace RE and full optimality with (i) adaptive learning and (ii) bounded optimality, based on (iii) the Lagrangian approach.

- The FOCs from the Lagrangian are

$$\begin{aligned} u_t &= -Q^{-1}W'x_t + (\beta/2)Q^{-1}B'E_t\lambda_{t+1} \\ \lambda_t &= -2Rx_t - 2Wu_t + \beta A'E_t\lambda_{t+1}, \end{aligned}$$

  where λ_t is the vector of shadow-prices of the state variables. These, the transition equation and the TVC identify optimal decision-making.

- Assuming adaptive learning we replace (A, B) with (A_t, B_t), estimated and updated by RLS. Under bounded rationality we replace λ_t and $E_t\lambda_{t+1}$ with $\hat{E}_t\lambda_t^*$ and $\hat{E}_t\lambda_{t+1}^*$.

$$\begin{aligned} u_t &= -Q^{-1}W'x_t + (\beta/2)Q^{-1}B_t'\hat{E}_t\lambda_{t+1}^* \\ \hat{E}_t\lambda_t^* &= -2Rx_t - 2Wu_t + \beta A_t'\hat{E}_t\lambda_{t+1}^*. \end{aligned}$$

  Thus: (1) given estimates x_t, B_t and $\hat{E}_t\lambda_{t+1}^*$, agents know how to choose their control u_t, (2) given x_t, u_t, A_t and $\hat{E}_t\lambda_{t+1}^*$, agents know how to revise their estimate $\hat{E}_t\lambda_t^*$ of the value of a unit of x_t today.

- We must also specify how agents forecast λ^*_{t+1}. We assume agents use the PLM

$$\lambda^*_t = Hx_t + \mu_t.$$

Agents do not know H, and at t use RLS to update their estimate to H_t, using a regression of $\hat{E}_s\lambda^*_s$ on x_s with data $s = 1, \ldots, t-1$. Then

$$\hat{E}_t\lambda^*_{t+1} = H_t(A_tx_t + B_tu_t).$$

- These equations + RLS defines SP-learning as a recursive system:

– Given estimates (A_t, B_t, H_t) and x_t, the u_t and $\hat{E}_t\lambda^*_{t+1}$ equations determine their values simultaneously.

– The transition equation $x_{t+1} = Ax_t + Bu_t + C\varepsilon_{t+1}$ gives x_{t+1}.

– Using data for x_{t+1}, x_t, u_t and $\hat{E}_t\lambda^*_t$ (from the $\hat{E}_t\lambda^*_t$ equation), estimates are updated using RLS to $(A_{t+1}, B_{t+1}, H_{t+1})$

The system can be written recursively as

$$
\begin{aligned}
R_t &= R_{t-1} + t^{-1}\left(x_{t-1}x_{t-1}' - R_{t-1}\right) \\
H_t &= H_{t-1} + t^{-1}R_t^{-1}x_{t-1}\left(E_{t-1}^*\lambda_{t-1}^* - H_{t-1}x_{t-1}\right)' \\
\hat{R}_t &= \hat{R}_{t-1} + \frac{1}{t}\left(\begin{pmatrix} x_{t-2} \\ u_{t-2} \end{pmatrix}\begin{pmatrix} x_{t-2}', & u_{t-2}' \end{pmatrix} - \hat{R}_{t-1}\right) \\
\begin{pmatrix} A_t \\ B_t \end{pmatrix} &= \begin{pmatrix} A_{t-1} \\ B_{t-1} \end{pmatrix} \\
&\quad + \frac{1}{t}\hat{R}_t^{-1}\begin{pmatrix} x_{t-2} \\ u_{t-2} \end{pmatrix}\left(x_{t-1} - \begin{pmatrix} A_{t-1}', & B_{t-1}' \end{pmatrix}\begin{pmatrix} x_{t-2} \\ u_{t-2} \end{pmatrix}\right)' \\
x_t &= Ax_{t-1} + Bu_{t-1} + C\varepsilon_t \\
u_t &= F(H_t, A_t, B_t)x_t \\
E_t^*\lambda_t^* &= \hat{T}(H_t, A_t, B_t)x_t \\
F(H_t, A_t, B_t) &= (2Q - \beta B_t'H_tB_t)^{-1}(\beta B_t'H_tA_t - 2W') \\
\hat{T}(H_t, A_t, B_t) &= -2R - 2WF(H_t, A_t, B_t) \\
&\quad + \beta A_t'H_t\left(A_t + B_tF(H_t, A_t, B_t)\right)
\end{aligned}
$$

Theorem 2 *Under standard assumptions, and assuming a suitable projection facility, under SP-learning* (H_t, A_t, B_t) *converges to* $(\bar{H}, A, B)$ *almost surely, where* $\bar{H} = \hat{T}(\bar{H}, A, B)$.

The heart of the argument focuses on $\hat{T}(H, A_t, B_t)$, which for given A_t, B_t maps the perceived shadow price parameters to the realized shadow-price parameters,

$$\hat{E}_t \lambda_t^* = \hat{T}(H_t, A_t, B_t) x_t.$$

We show that (as with E-stability under LS learning) convergence is determined by stability of

$$dH/d\tau = \hat{T}(H, A, B) - H,$$

and we then to show that this differential equation is locally stable at the fixed point $\bar{H}$.

Theorem 2 is a striking result:

- Decisions converge asymptotically to fully rational forecasts and fully optimal decisions.

- By estimating 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.

Euler-equation learning

- The paper discusses the relationship between SP-learning and Euler-equation learning.

- Euler equations are traditionally derived from variational arguments and give FOCs that do not depend on Lagrange multipliers.

- First-order (one-step-ahead) Euler equations exist only in special cases: (1) $\mathsf{dim}(u) \geq \mathsf{dim}(x)$ and $\mathsf{det}\, g_u \neq 0$ or (2) $g_x = 0$.

- Higher-order Euler equations exist more generally.

- Proposition 3 shows that for the LQ problem with $\dim(u) = \dim(x)$ **and** $g_x = 0$ then suitably specified EE-learning is equivalent to SP-learning (in the sense that the largest eigenvalue of the $\hat{T}-$map, which governs asymptotic speed of convergence under learning, is the same).

- In general SP-learning and EE-learning are **not** equivalent. We give an example below.

Example: SP Learning in a Crusoe economy

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

$$\text{s.t.} \quad s_t = A_1 s_{t-1} + A_2 s_{t-2} - c_t + \mu_{t+1}$$

Output is fruit/sprouting trees. Under SP-learning Bob estimates the SPs of new and old trees:

$$\lambda_{it}^* = a_{it} + b_{it} s_{t-1} + d_{it} s_{t-2}, \text{ for } i = 1, 2, \text{ and thus}$$

$$\hat{E}_t \lambda_{it+1}^* = a_{it} + b_{it}(A_{1t-1} s_{t-1} + A_{2t-1} s_{t-2} - c_t) + d_{it} s_{t-2}, \text{ for } i = 1, 2.$$

These plus the FOC for the control

$$c_t = \hat{b} - 0.5 \beta \hat{E}_t \lambda_{1t+1}^*.$$

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

The FOCs for the states give updated estimates of SPs

$$\begin{aligned}\hat{E}_t\lambda^*_{1t} &= -2\phi s_{t-1} + \beta A_{1t} E_t \lambda^*_{1t+1} + \beta E_t \lambda^*_{2t+1} \\ \hat{E}_t\lambda^*_{2t} &= \beta A_{2t} E_t \lambda^*_{1t+1},\end{aligned}$$

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

For this example EE-learning is also possible (by substituting out the SPs)

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

To implement EE-learning agents forecast using estimates of

$$c_t = a_3 + b_3 s_{t-1} + d_3 s_{t-2}.$$

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 1.

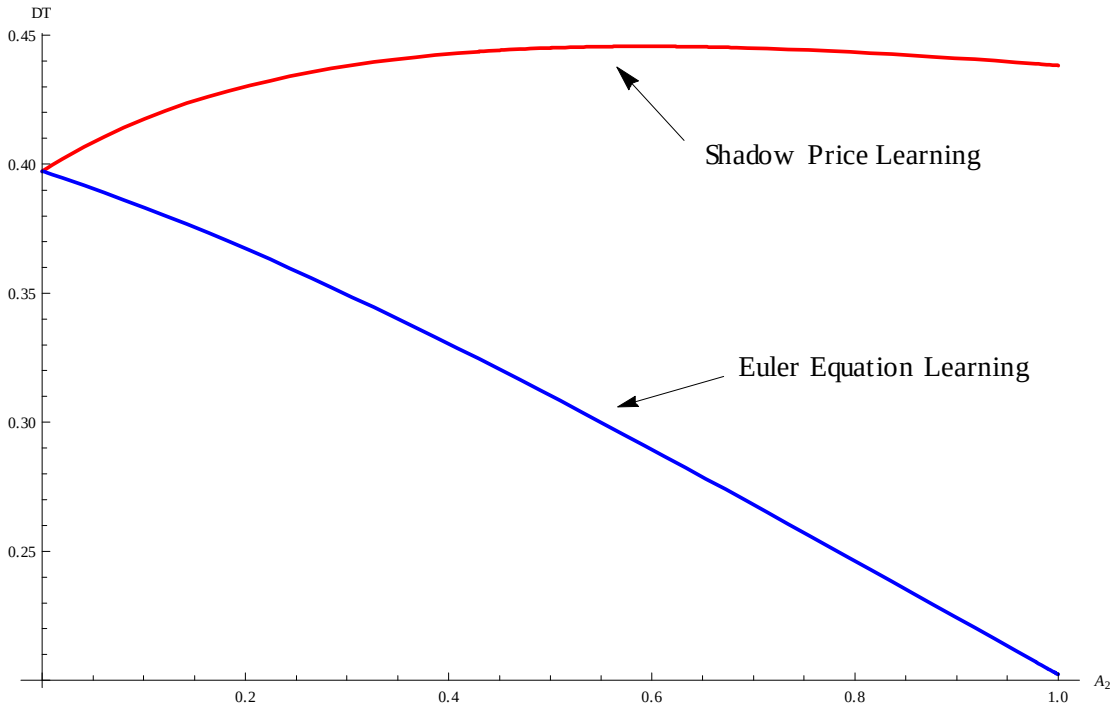

Figure 1: Largent eigenvalue of $DT_H(H, A, B)$ under SP and EE learning.

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 info 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.

Example: SP Learning in a Ramsey Model

Our formal results for SP learning are for LQ models. Most DSGE models use a more general setting. However we earlier described how SP learning can be applied to a more general nonlinear setting, in which agents use linear forecasting techniques.

We illustrate with the stochastic Ramsey model. Households have one unit of labor and maximize

$$\begin{aligned} \max \quad & E \sum \beta^t u(c_t) \\ & s_t = (1 + r_t) s_{t-1} + w_t - c_t. \end{aligned}$$

Competitive firms use CRTS technology and $y = zf(k)$ where y, k are output, capital per unit of labor and $\log z_t$ is AR(1) stationary with $Ez_t = 1$.

Households will choose

$$u'(c_t) = \beta \hat{E}_t \lambda^*_{t+1},$$

where λ^*_t is the shadow value of an additional unit of s_{t-1}, which here is equal to $(1 + r_t)u'(c_t)$. We assume agents estimate

$$\lambda^*_t = a + bk_t + ez_t$$

and use this and the linearized capital accumulation equation to forecast λ^*_{t+1}.

Under SP-learning the recursive system is

$$\begin{aligned}
z_t &= \varepsilon_t z^{\rho}_{t-1} \\
c_t &= c(k_t, z_t, \phi_{t-1}), \text{ where } \phi_t = (a_t, b_t, e_t)' \\
\lambda^*_t &= (1 + z_t f'(k_t) - \delta)u'(c_t) \\
R_t &= R_{t-1} + \gamma_t(x_t x'_t - R_{t-1}) \\
\phi_t &= \phi_{t-1} + \gamma_t R^{-1}_t x_t(\lambda^*_t - \phi'_{t-1} x_t) \\
k_{t+1} &= z_t f(k_t) + (1 - \delta)k_t - c_t.
\end{aligned}$$

Illustration

For log utility, Cobb-Douglas production, and $\delta = 1$, we can obtain the explicit RE solution and analytical REE shadow price λ_t function.

The red line is initial beliefs. Under learning their is convergence to the black line. The dashed blue line is the $\lambda(k)$ in the REE.

Long-run beliefs correspond to first order to the true dependence in the REE.

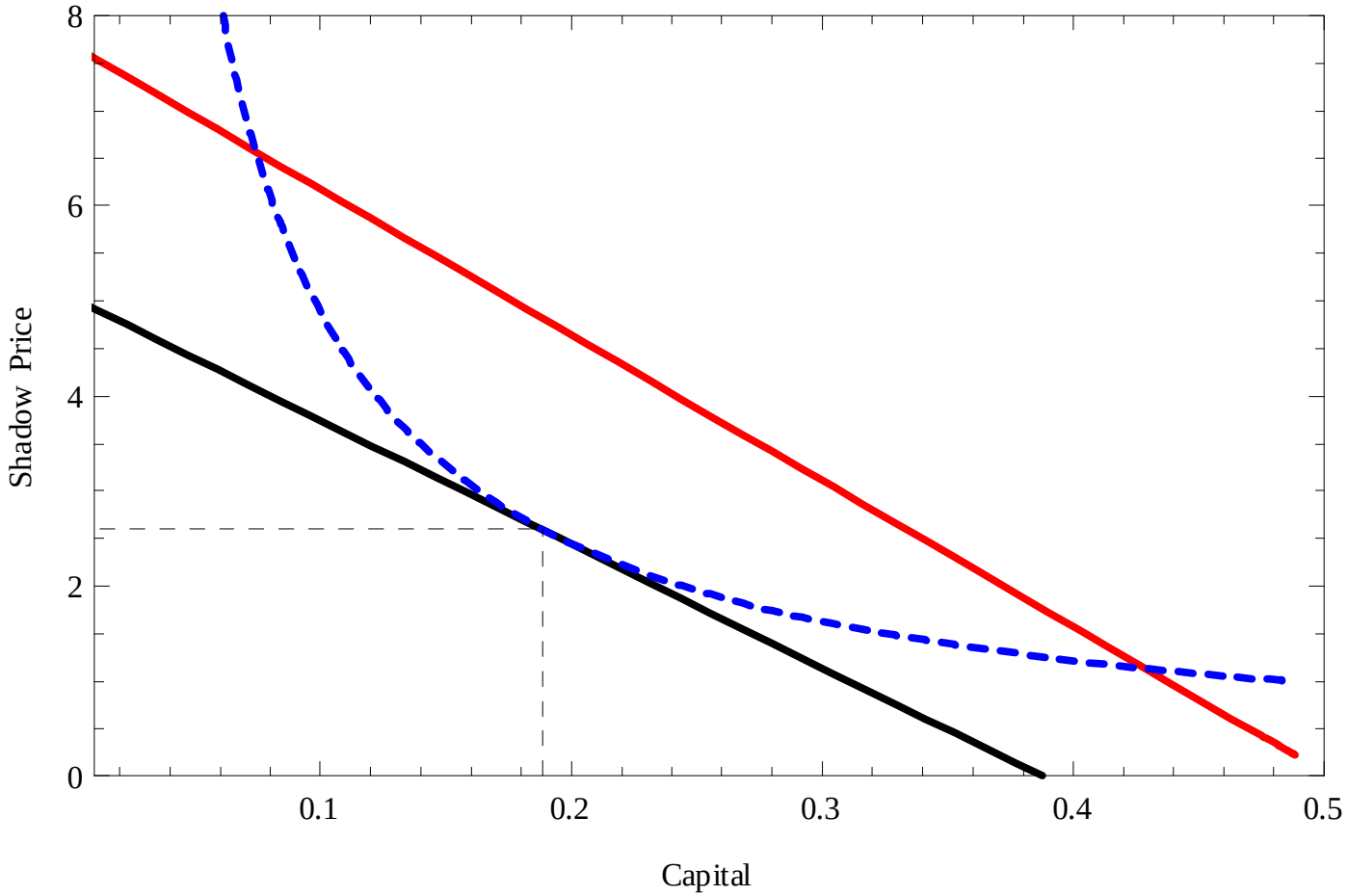

Figure 2: Red: initial beliefs. Black: final beliefs. Blue: true $\lambda(k)$ in REE.

Conclusions

- SP learning can be applied to more general set-ups and in general equilibrium models.

- In special cases SP-learning reduces to Euler-equation learning, but SP-learning is more general.

- Advantage of SP-learning: agents need only solve 2-period optimization problems using one-step ahead forecasts of states and shadow prices.

- SP learning is boundedly optimal but is also asymptotically optimal.

- As a model of bounded rationality in a dynamic, stochastic setting, SP-learning has the advantage of simplicity, generality and economic intuition.

- Future work includes:

  – Application of SP learning in more sophisticated models.

  – Careful analysis of the relationships between SP-learning and alternative learning rules.

  – As with expectations, persistent deviations from full optimization may be natural to consider and of interest to explore.