Exploring Bayesian Active Learning of Drifting Coefficient Regressions

Sergei Morozov

Morgan Stanley

Far Eastern Meeting of Econometric Society, 2009

Talk Outline

Exploring Bayesian Active Learning of Drifting Coefficient Regressions

- Setup
  - Bayesian active learning control: what, why and how
  - Drifting parameter models
- Contribution
  - Explore quality of larger variety of suboptimal solutions
  - New bounds for actively optimal solution
  - New features of actively optimal solution
- Conclusions and related work

Bayesian Active Learning Control

What is Bayesian Active Learning Control?

- Stabilizing discounted deviations from desired time path
- Imperfect information setting (model uncertainty)
- Bayesian updating of model uncertainties
- Active Learning control is a compromise between stabilization now and experimentation in order to learn to stabilize better in the future
- Information accumulation is endogenous and factors in decision making

Bayesian Active Learning Control

Why Study Bayesian Active Learning Control?

- Consumption smoothing
- Macroeconomic stabilization under model uncertainty
  - Monetary and fiscal policies under model uncertainty
  - Exchange rate targeting
  - Discovery of good policies
  - Measurement of trade-off possibilities among competing aims
  - Evaluation of policy proposals
  - Appraisal of historical policies
  - Selection of sets of policy instruments based on their effectiveness
- Resource control
- Monopolistic pricing with unknown demand
- New product introductions
- Price discovery in financial asset pricing

Bayesian Active Learning Control

How to Study Bayesian Active Learning Control?

- Solve imperfect information problem by dynamic programming – *Cursed by Curse of Dimensionality*
- Find good approximating suboptimal solution by opening Pandora box of bounded rationality
- Study suboptimal solutions in order to:
  - estimate the size of experimentation component of policy in terms of policy, loss function and outcomes
  - appreciate departure into bounded rationality
  - provide good starting points for active learning optimal algorithms
  - provide heuristics for cases where active optimal policy is too hard to find

Drifting Parameter Models

Nine out of ten people who change their minds are wrong the second time too.

–Anonymous

- Encapsulates idea of continuously adapting economic environment
- Captures lack of consensus about stability of data generating process over time
- Empirical justifications:
  - Canova (2006): lack of posterior tightening in the small-scale New Keynesian model of US economy
  - Cogley-Sargent (2001): departures from time-invariance in the US inflation dynamics
  - etc. ...

Control of Drifting Coefficient Regression

$$\min_{\{u_t\}_{t=0}^{\infty}} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \delta^t \left((x_t - x^*)^2 + \omega (u_t - u^*)^2 \right) \right]$$

subject to

$$\begin{aligned}
\text{observed state: } x_t &= \alpha + \beta_t u_t + \gamma x_{t-1} + \epsilon_t, \\
\text{random coefficients: } \beta_t &= \beta_{t-1} + \eta_t. \\
\begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix} &\sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \sigma_\epsilon^2 & 0 \\ 0 & \sigma_\eta^2 \end{bmatrix} \right), \\
\alpha, \gamma &- \text{known}
\end{aligned}$$

Bayesian Learning Dynamics (Kalman filtering): $\beta_t \sim \mathcal{N}(\mu_t, \Sigma_t)$

$$\begin{aligned}
\mu_t &= \mu_{t-1} + (\Sigma_{t-1} + \sigma_\eta^2) u_t \left(u_t (\Sigma_{t-1} + \sigma_\eta^2) u_t + \sigma_\epsilon^2 \right)^{-1} (x_t - \alpha - \mu_{t-1} u_t - \gamma x_{t-1}) \\
\Sigma_t &= \Sigma_{t-1} + \sigma_\eta^2 - (\Sigma_{t-1} + \sigma_\eta^2)^2 u_t^2 \left(u_t^2 (\Sigma_{t-1} + \sigma_\eta^2) + \sigma_\epsilon^2 \right)^{-1}
\end{aligned}$$

- Variance updating is deterministic
- Decreasing variance if $\sigma_\eta^2 = 0$, so that learning would converge
- Experimentation never ceases if $\sigma_\eta^2 > 0$
- First solved in Beck and Wieland (2002) by numerical dynamic programming
- Our approach is a refinement

Actively Optimal Control

Dynamic Programming Formulation

- Information state $\mathcal{I} = \left\{ \left(\mu_{t+1|t}, \Sigma_{t+1|t}\right)' \right\}$
- Extended state $\mathcal{S} = \mathcal{X} \times \mathcal{I} \subseteq \mathbb{R}^3$
- Bayesian updating induces nonlinear mapping on information state $B(\cdot, x_{t-1}, u_t) : \mathcal{S} \longrightarrow \mathcal{S}$
- Stationary Bellman Equation for continuation value (cost-to-go)

$$
\begin{aligned}
V(S_t) = \min_{\{u_{t+1}\}} \Big\{ & L(S_t, u_{t+1}) \\
& + \delta \int V\left(B(S_t, \alpha + \beta_{t+1} u_{t+1} + \gamma x_t + \epsilon_{t+1}, u_{t+1})\right) p(\beta_{t+1}|S_t) q(\epsilon_{t+1}) d\beta_{t+1} d\epsilon_{t+1} \Big\} \\
=: \; & T[V](S_t),
\end{aligned}
$$

where $L(S_t, u_{t+1})$ is expected one-period loss

$$
L(S_t, u_{t+1}) = \int \left((\alpha + \beta_{t+1} u_{t+1} + \gamma x_t + \epsilon_{t+1} - x^*)^2 + \omega (u_{t+1} - u^*)^2 \right) p(\beta_{t+1}|S_t) q(\epsilon_{t+1}) d\beta_{t+1} d\epsilon_{t+1},
$$

and $T[V]$ is Bellman functional operator

- T is a contraction mapping, value function iterations converge (Kiefer-Nyarko, 1989)
- Value function may have kinks and policy function may have discontinuities
- Policy iteration used to reduce computing time
- Adaptive space discretization - finer mesh near policy discontinuities or areas of high curvature

Actively Optimal Control

Useful Analytic Bounds

- There are policies whose Q-factors could be computed analytically
  - Do-nothing policy: $u \equiv 0$

$$V_t^0(x_t) = \frac{(\alpha + x_t - x^*)^2 - \delta\gamma\left((x^*)^2 - \alpha^2 - \gamma x^*(2\alpha - x^*) + \gamma x_t^2(1+\gamma) - 2x_t(x^* - \alpha + \gamma x^*)\right)}{(1-\delta)(1-\gamma\delta)(1-\gamma^2\delta)}$$
$$+ \frac{\gamma^3\delta^2\,(x_t - x^*)^2}{(1-\delta)(1-\gamma\delta)(1-\gamma^2\delta)} + \frac{\sigma_\epsilon^2}{(1-\delta)(1-\gamma^2\delta)} + \frac{\omega\,(u^*)^2}{1-\delta}.$$

  - Pseudo-myopic policy

$$u_{t+1}^{pm} = \arg\min \mathbb{E}_t\left\{L(x_{t+1}, u_{t+1}) + \delta V^0(x_{t+1})\right\}.$$

    has analytic expression for $V_t^{pm}(x_t, \mu_t, \Sigma_{t|t})$ as well

- Analytic Q-factors bound actively optimal cost-to-go but can be translated into policy space

$$\mathbb{E}_t\left[L(x_{t+1}, u_{t+1}^*)\right] \leq V^* \leq \min\left\{V_t^0(x_t), V_t^{pm}(x_t, \mu_t, \Sigma_{t|t})\right\}$$

- Bounds are somewhat loose, especially in extreme regions
- Useful to prove validity of using bounded state-space algorithm with reflective barriers to approximate solution on the unbounded domain
- Useful to provide bounds to minimization step and for initial guesses

Actively Optimal Dual Control

Properties

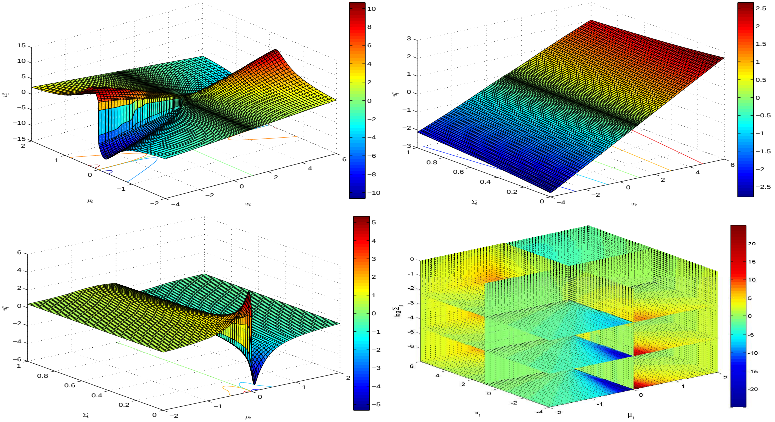


Figure: 1. Actively adaptive optimal control. Parameters: $\alpha = \omega = u^* = 0$, $\gamma = 0.9$, $\delta = 0.75$, $x^* = 1$, $\sigma_\epsilon^2 = 1.0$, $\sigma_\eta^2 = 0.01$.

- Aggressive experimentation in the vicinity of x^*
- Possible discontinuity at x^* due to nonconvexity

Certainty Equivalent Control

Dynamic Programming Formulation

- Solves infinite horizon LQG problem with constant coefficients, i.e. treating β_{t+1} as known constant $\beta_{t+1} = \mu_{t+1|t} = \mu_t$
- Bellman equation

$$V(x_t) = \min_{\{u_{t+1}\}} \Big\{ \mathbb{E}_t \left(\alpha + \beta_{t+1} u_{t+1} + \gamma x_t + \epsilon_{t+1} - x^*\right)^2 + \omega (u_{t+1} - u^*)^2 + \delta \mathbb{E}_t V(\alpha + \beta_{t+1} u_{t+1} + \gamma x_t + \epsilon_{t+1}) \Big\}$$

  has a solution that is quadratic in x_t, $V^{CE}(x_t) = A x_t^2 + 2B x_t + C$
- CE Policy is linear in x_t

$$u_{t+1}^{CE} = -\frac{\mu_t \gamma (1 + \delta A)}{\mu_t^2 + \omega + \delta A \mu_t^2} x_t + \frac{\mu_t (x^* - \alpha) + \omega u^* - \delta \mu_t B - \delta \alpha \mu_t A}{\mu_t^2 + \omega + \delta A \mu_t^2}$$

- Coefficient $A > 0$ solves univariate version of algebraic Riccati equation

$$\gamma^2 \omega (1 + \delta A) - A \left(\mu_t^2 (1 + \delta A) + \omega \right) = 0$$

- $$B = \frac{\omega \left(\gamma (\mu_t u^* + \alpha) + \mu_t \gamma \delta A u^* - \gamma x^* \right)}{\mu_t^2 + \omega + \delta A \mu_t^2 - \gamma \delta \omega}$$

Certainty Equivalent Control

Properties

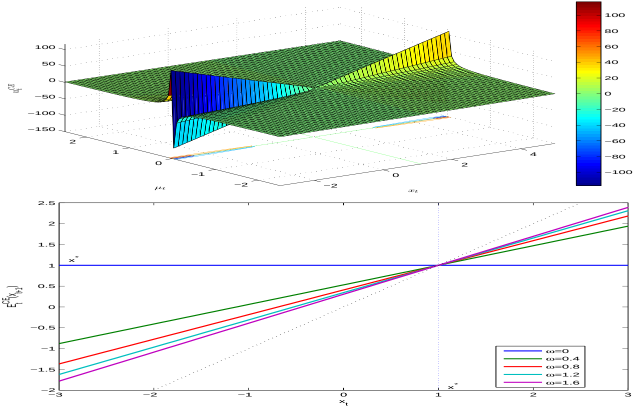


Figure: 2. Certainty equivalent control. Parameters: $\alpha = \omega = u^* = 0$, $\gamma = 0.9$, $\delta = 0.75$, $x^* = 1$.

- Aggressive experimentation in the vicinity of x^*
- $\omega = 0$ implies instantaneous adjustment to target $\mathbb{E}_t^{CE}(x_{t+1}) = x^*$

Anticipated Utility Control

Dynamic Programming Formulation

- Solves infinite horizon LQG problem with constant but random coefficients while ignoring future learning (Bayesian linear regulator)
- Bellman equation

$$V(x_t) = \min_{u_{t+1}} \{\mathbb{E}_t \left(\alpha + \beta_{t+1} u_{t+1} + \gamma x_t + \epsilon_{t+1} - x^*\right)^2 + \omega(u_{t+1} - u^*)^2$$
$$+ \delta \mathbb{E}_t V(\alpha + \beta_{t+1} u_{t+1} + \gamma x_t + \epsilon_{t+1})\},$$

$\beta_{t+1} \sim \mathcal{N}\left(\mu_{t+1|t}, \Sigma_{t|t} + \sigma_\eta^2\right)$, has a solution that is quadratic in x_t, $V^{AE}(x_t) = Ax_t^2 + 2Bx_t + C$

- AU Policy is linear in x_t

$$u_{t+1}^{AU} = -\frac{\gamma(1+\delta A)\mu_t}{(\Sigma_t + \sigma_\eta^2 + \mu_t^2)(1+\delta A) + \omega} x_t + \frac{(x^* - \alpha(1+\delta A) - \delta B)\mu_t + \omega u^*}{(\Sigma_t + \sigma_\eta^2 + \mu_t^2)(1+\delta A) + \omega}$$

- Coefficient A solves univariate version of algebraic Riccati equation

$$-\frac{-\gamma^2 \mu_t^2 (1+\delta A)^2}{(\Sigma_t + \sigma_\eta^2 + \mu_t^2)(1+\delta A) + \omega} + \gamma^2(1+\delta A) = A$$

- B solves $\frac{\gamma\mu_t(1+\delta A)\left(\mu_t(x^* - \alpha(1+\delta A) - \delta B) + \omega u^*\right)}{(\Sigma_t + \sigma_\eta^2 + \mu_t^2)(1+\delta A) + \omega} + \gamma(\alpha - x^*) + \delta\alpha\gamma A = (1 - \delta\gamma)B$

Anticipated Utility Control

Properties

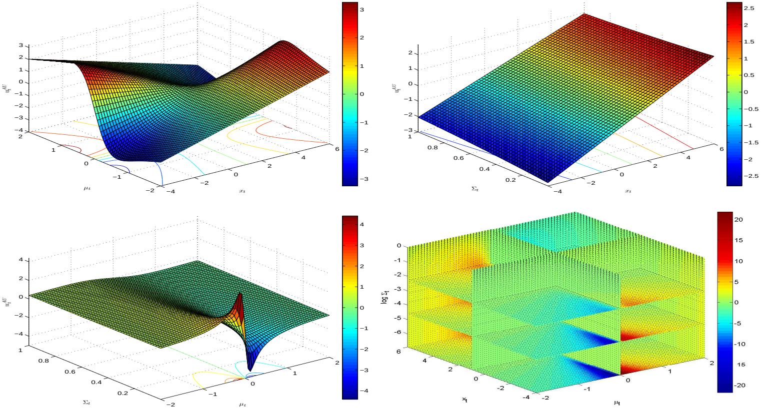


Figure: 3. Anticipated utility control. Parameters: $\alpha = \omega = u^* = 0$, $\gamma = 0.9$, $\delta = 0.75$, $x^* = 1$, $\sigma_\epsilon^2 = 1.0$, $\sigma_\eta^2 = 0.01$.

- Policy smoothing due to caution

Markov Jump Linear Quadratic Control

Formulation

- Solves infinite horizon LQG problem with random coefficients that follow unobserved finite state Markov chain
- MJLQ state equation

$$X_{t+1} = A_{s(t+1)}X_t + B_{s(t+1)}U_{t+1} + C_{s(t+1)}\epsilon_{t+1},$$

$s(t+1) \in \{1, \dots, S\},$, $p_{t+1,j} = \Pr(s(t+1) = j)$, $p_{t+1} = \left(p_{t+1,1}, \dots, p_{t+1,S}\right)' = P'p_t$,
$P_{ij} = \Pr\{s(t+1) = j|s(t) = i\}$, $i, j = 1, \dots, S$.

- MJLQ Bellman equation

$$\begin{aligned} V(X_t, p_{t+1}) &= X_t' W(p_{t+1})X_t + w(p_{t+1}) = \\ &= \min_{U_{t+1}} \left\{ X_t' Q X_t + U_{t+1}' R U_{t+1} + \delta \mathbb{E}_t V(X_{t+1}, p_{t+1}) \right\} \\ &= \min_{U_{t+1}} \left\{ X_t' Q X_t + U_{t+1}' R U_{t+1} \right. \\ &\left. + \delta \sum_{j,k} p_{t+1,j} P_{jk} \left(X_{t+1,k}' W(p_{t+2}) X_{t+1,k} + w(p_{t+2}) \right) \right\}, \end{aligned}$$

Markov Jump Linear Quadratic Control

Solution

- $$U_{t+1} = -G(p_{t+1})^{-1}K(p_{t+1})X_t,$$

  where

  $$G(p_{t+1}) = R + \delta \sum_{j,k} p_{t+1,j} P_{jk} B_k' W(P' p_{t+1}) B_k,$$

  $$K(p_{t+1}) = \delta \sum_{j,k} p_{t+1,j} P_{jk} B_k' W(P' p_{t+1}) A_k.$$

- Matrix $W(p_t)$ solves Riccati equation

  $$W(p_{t+1}) = Q + \delta \sum_{j,k} p_{t+1,j} P_{jk} A_k' W(P' p_{t+1}) A_k - K(p_{t+1})' G(p_{t+1})^{-1} K(p_{t+1}).$$

- Riccati equation can be solved by receding control
  - set distant horizon with terminal cost-to-go from the problem with observed regimes
  - iterate backwards to obtain $W(p_{t+1})$
  - expand horizon until convergence
  - observed regime solution satisfies system of coupled Riccati equations, which could be uncoupled by change of variables
  - uncoupled Riccati equations could be solved by doubling algorithm (`doubleo.m` and `olrp.m`)
- $W(p_{t+1})$ for the entire simplex of probabilities requires function of approximation methods

Markov Jump Linear Quadratic Control

Approximating Continuous Drift by Markov Jump Linear Quadratic Model

- Partition the support of $\mathcal{N}(\mu_{t+1|t}, \Sigma_{t+1|t})$ distribution into S segments of equal probability, and define S states (regimes) as the expected values of respective truncated normal distributions over each segment.
- Define transition probability matrix by discretizing the probability distribution of β_{t+2} conditional on $\beta_{t+1} = \beta_k$ for each $k = 1, \ldots, S$.
- Set $X_t = \begin{pmatrix} 1 \\ x_t - x^* \end{pmatrix}$, $U_t = u_t$,
  $A_k = \begin{pmatrix} 1 & 0 \\ \alpha + (\gamma - 1)x^* + \beta_k u^* & \gamma \end{pmatrix}$,
  $B_k = \begin{pmatrix} 0 \\ \beta_k \end{pmatrix}$,
  $C_k = \begin{pmatrix} 0 \\ \sigma_\epsilon \end{pmatrix}$,
  $Q = \begin{pmatrix} \gamma^2(x^*)^2 + (\alpha - x^*)^2 + 2\gamma(\alpha - x^*)x^* + \omega(u^*)^2 + \sigma_\epsilon^2 & \gamma^2 x^* + \gamma(\alpha - x^*) \\ \gamma^2 x^* + \gamma(\alpha - x^*) & \gamma^2 \end{pmatrix}$,
  $N = \begin{pmatrix} (\alpha + (\gamma - 1)x^*) - \omega u^*) \left(\sum_{j=1}^S p_{j,t+1}\beta_j\right) \\ \gamma \sum_{j=1}^S p_{j,t+1}\beta_j \end{pmatrix}$,
  $R = \omega + \sum_{j=1}^S p_{j,t+1}\beta_j^2$.

Markov Jump Linear Quadratic Control

Properties

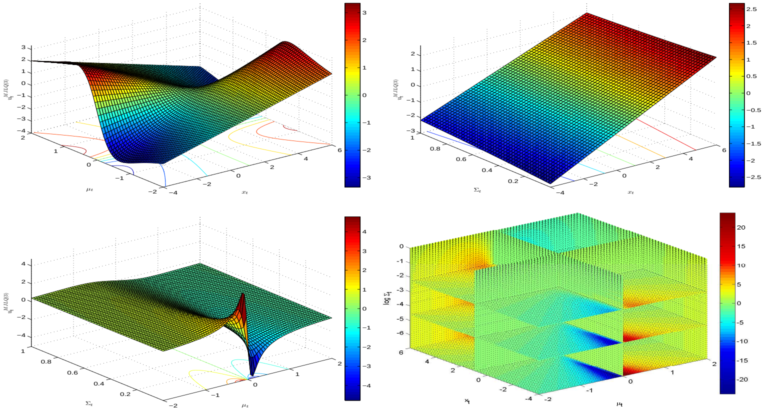


Figure: 4. MJLQ(3) control. Parameters: $\alpha = \omega = u^* = 0$, $\gamma = 0.9$, $\delta = 0.75$, $x^* = 1$, $\sigma_\epsilon^2 = 1.0$, $\sigma_\eta^2 = 0.01$.

- Policy smoothing due to caution
- Similar to anticipated utility

Limited Lookahead Control

Definition

n-period lookahead policy $u_{t+1}^{l(n)}$ is part of the solution to the following finite-horizon problem:

$$\min_{u_{t+1},\dots,u_{t+n+1}} \Bigg\{ \mathbb{E}_t \left[\left(\alpha + \beta_{t+1} u_{t+1} + \gamma x_t + \epsilon_{t+1} - x^* \right)^2 \right] + \sum_{\tau=1}^{n} \delta^\tau \mathbb{E}_t \left[(x_{t+\tau+1} - x^*)^2 \middle| u_{t+1}, \dots, u_{t+\tau} \right] + \omega \sum_{\tau=0}^{n} \delta^\tau (u_{t+\tau+1} - u^*)^2 \Bigg\}$$

where

$$\mathbb{E}_t \left[(x_{t+\tau+1} - x^*)^2 \middle| u_{t+1}, \dots, u_{t+\tau} \right] = \left(\alpha \frac{1-\gamma^{\tau+1}}{1-\gamma} + \gamma^{\tau+1} x_t - x^* \right)^2 + \sum_{s=0}^{\tau} \gamma^{2s} \sigma_\epsilon^2 + 2 \left(\alpha \frac{1-\gamma^{\tau+1}}{1-\gamma} + \gamma^{\tau+1} x_t - x^* \right) \mu_t \sum_{s=0}^{\tau} \gamma^s u_{t+\tau+1-s} + \mathbb{E}_t \left[\left(\sum_{s=0}^{\tau} \gamma^s \beta_{t+\tau+1-s} u_{t+\tau+1-s} \right)^2 \middle| u_{t+1}, \dots, u_{t+\tau} \right].$$

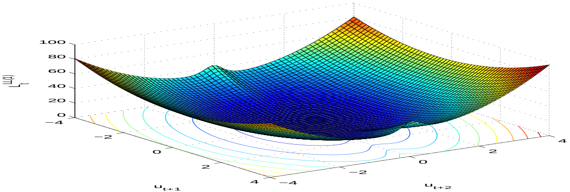
- We need to compute all cross-moments $\mathbb{E}_t \left(\beta_{t+j} \beta_{t+k} \middle| u_{t+1}, \dots, u_{t+\tau} \right)$ for all $j, k = 1, \dots, \tau + 1$ unless $\gamma = 0$.
- Use fixed point Kalman smoother with auxiliary constant state to compute all cross-covariances
- Solving for n-period limited lookahead control is equivalent to solving polynomial equation of degree $4n + 1$

One-Period Lookahead Control

Formulation

$$\min_{u_{t+1},u_{t+2}} \mathbb{E}_t \Big\{ (\alpha + \beta_{t+1} u_{t+1} + \gamma x_t + \epsilon_{t+1} - x^*)^2 + \omega (u_{t+1} - u^*)^2$$
$$+ \delta \Big[(\alpha + \beta_{t+2} u_{t+2} + \gamma x_{t+1} + \epsilon_{t+2} - x^*)^2 + \omega (u_{t+2} - u^*)^2 \Big] \Big\}.$$

$$\begin{aligned}
\mathbb{E}_t \beta_{t+1}^2 &= \mu_t^2 + \Sigma_t + \sigma_\eta^2, \\
\mathbb{E}_t \left[\beta_{t+1}^2 | u_{t+1} \right] &= \mu_t^2 + \Sigma_{t+1}(u_{t+1}), \\
\mathbb{E}_t \left[\beta_{t+1}\beta_{t+2} | u_{t+1} \right] &= \mu_t^2 + \Sigma_{t+1}(u_{t+1}), \\
\mathbb{E}_t \left[\beta_{t+2}^2 | u_{t+1} \right] &= \mu_t^2 + \Sigma_{t+1}(u_{t+1}) + \sigma_\eta^2,
\end{aligned}$$



1-period lookahead loss

- Non-convexities can develop similar to "unlimited" lookahead

One-Period Lookahead Control

Properties

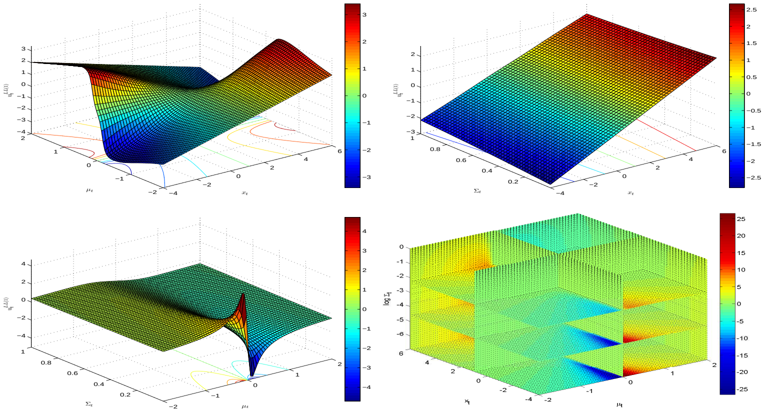


Figure: 5. One-period lookahead control. Parameters: $\alpha = \omega = u^* = 0$, $\gamma = 0.9$, $\delta = 0.75$, $x^* = 1$, $\sigma_\epsilon^2 = 1.0$, $\sigma_\eta^2 = 0.01$.

- Policy smoothing due to caution
- Slightly more nonlinear

Passively Optimal Control

Dynamic Programming Formulation

- Solves infinite horizon LQG problem with random coefficients that follow random walk but ignoring future learning
- Bellman equation

$$V(x_t, \mu_{t+1|t}, \Sigma_{t+1|t}) = \min_{\{u_{t+1}\}} \Big\{ L(x_t, \mu_{t+1|t}, \Sigma_{t+1|t}, u_{t+1})$$
$$+ \delta \int \int V\left(\alpha + \mu_{t+2|t} u_{t+1} + \gamma x_t + \epsilon_{t+1}, \mu_{t+2|t}, \Sigma_{t+2|t}\right)$$
$$\times\ p(\mu_{t+2|t}) q(\epsilon_{t+1}) d\mu_{t+2|t} d\epsilon_{t+1} \Big\}.$$

- Reduces to anticipated utility if $\sigma_\eta^2 = 0$
- Double integration makes solution more difficult to compute
- Unbounded drift of predictive variance in the absence of new observations requires receding finite absorbing boundary until no solution change

Passively Optimal Control

Properties

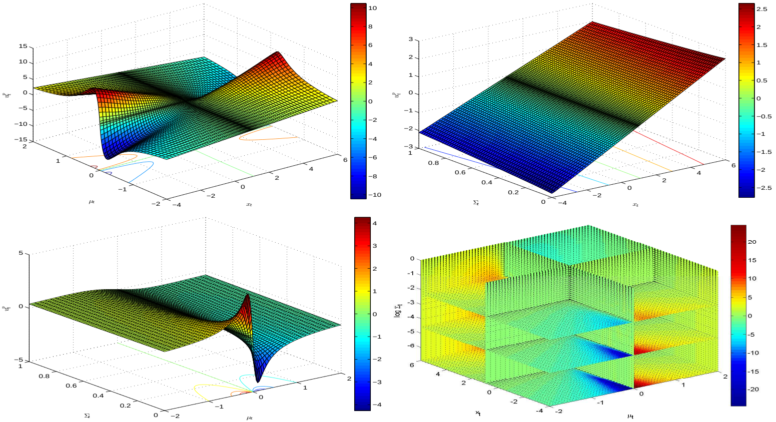
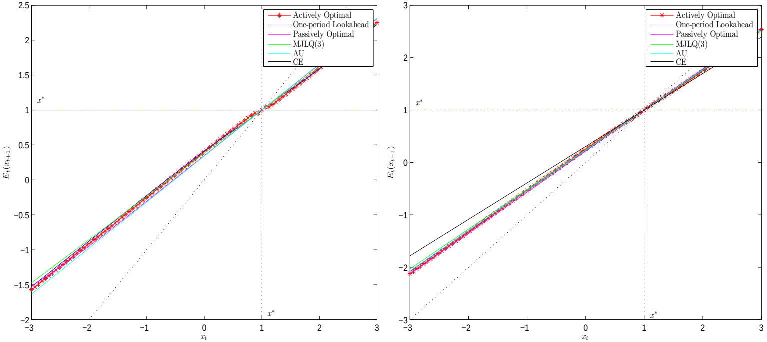


Figure: 6. Passively adaptive optimal control. Parameters: $\alpha = \omega = u^* = 0$, $\gamma = 0.9$, $\delta = 0.75$, $x^* = 1$, $\sigma_\epsilon^2 = 1.0$, $\sigma_\eta^2 = 0.01$.

- Policy smoothing due to caution

Gradualism

... a little stodginess at the central bank is entirely appropriate (A. Blinder, 1998)

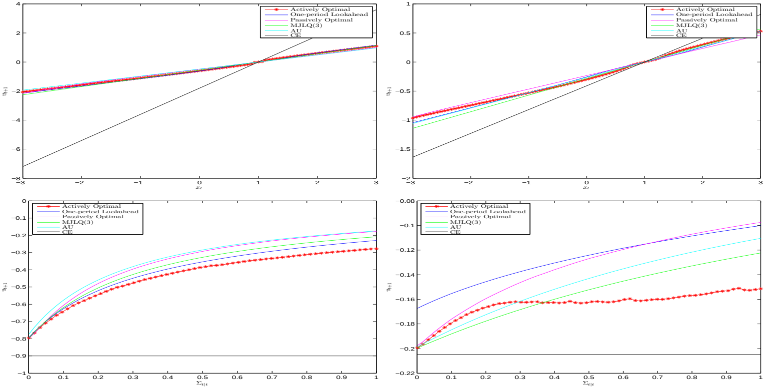


$\omega = 0$ $\omega = 1.6$

Figure: 7. Expected target state under alternative policies. State coordinates: $\mu_t = -0.5$, $\Sigma_t = 0.64$. Parameter values: $\alpha = 0.1$, $\gamma = 0.9$, $\delta = 0.75$, $x^* = 1$, $u^* = 0$, $\sigma_\epsilon^2 = \sigma_\eta^2 = 0.04$.

- Only CE policy with $\omega = 0$ implies one-step adjustment to target (in expectation)

Policy Function Comparison



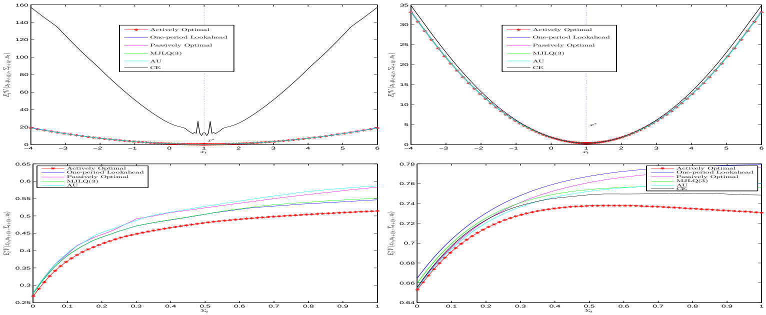
$\omega = 0$ $\omega = 1.6$;

Figure: 8. Policy functions under alternative policies. State coordinates for top plots: $\mu_t = -0.5$, $\Sigma_t = 0.64$. State coordinates for bottom plots: $\mu_t = -0.5$, $x_t = 0.5$. Parameter values: $\alpha = 0.1$, $\gamma = 0.9$, $\delta = 0.75$, $x^* = 1$, $u^* = 0$, $\sigma_\epsilon^2 = \sigma_\eta^2 = 0.04$.

- Differences among policies are most pronounced along parameter uncertainty dimension

Comparison of Expected Cost-to-go Functions

Optimal Q-factors take account of future learning under given policy rule, unlike objective functions of suboptimal policies



$\omega = 0$ $\omega = 1.6$

Figure: 9. Actively adaptive value function under alternative policies. State coordinates for top plots: $\mu_t = -0.5$, $\Sigma_t = 0.64$. State coordinates for bottom plots: $\mu_t = -0.5$, $x_t = 0.5$. Parameter values: $\alpha = 0.1$, $\gamma = 0.9$, $\delta = 0.75$, $x^* = 1$, $u^* = 0$, $\sigma_\epsilon^2 = \sigma_\eta^2 = 0.04$.

- Active advantage grows with parameter uncertainty

Optimal Q-factor of Anticipated Utility Policy

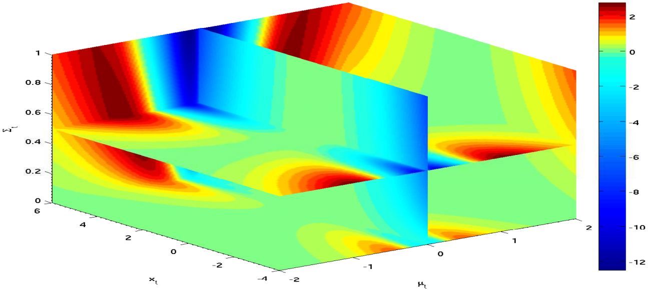


Figure: 10. Volumetric plot of $V^*(S, u^{AU}) - V^{AU}(S, u^{AU})$. Parameters: $\alpha = 0.1$, $\gamma = 0.9$, $\delta = 0.75$, $\omega = 1.6$, $x^* = 1$, $u^* = 0$, $\sigma_\epsilon^2 = \sigma_\eta^2 = 0.04$.

- $V^*(S, u^{AU}) - V^{AU}(S, u^{AU})$ could be negative or positive; using $V^{AU}(S, u^{AU})$ could distort inference of the benefit to intentional experimentation

Expected State Dynamics

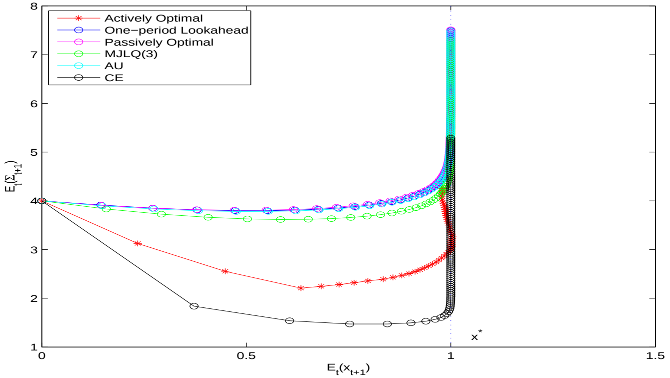


Figure: 11. Expected dynamics of extended state under different policies. Parameter values: $\alpha = 0.1$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 1.0$, $\sigma_\eta^2 = 0.04$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Mean belief: $\mu_t = 0.5$. Starting values: $x_t = 0$, $\Sigma_t = 4$.

- Actively optimal and certainty equivalent policies induce most learning via intentional and accidental experimentation

Simulated Controls

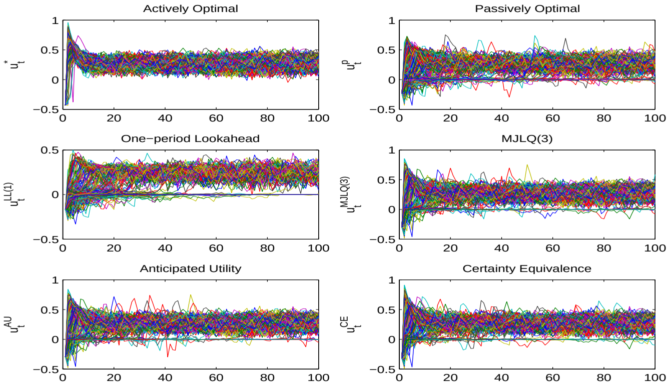


Figure: 12. Simulated multiple time-series of control under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$. Number of simulations: $NMC = 400$.

- Only actively optimal policy does not get stuck at $u_t = 0$

Simulated Physical States

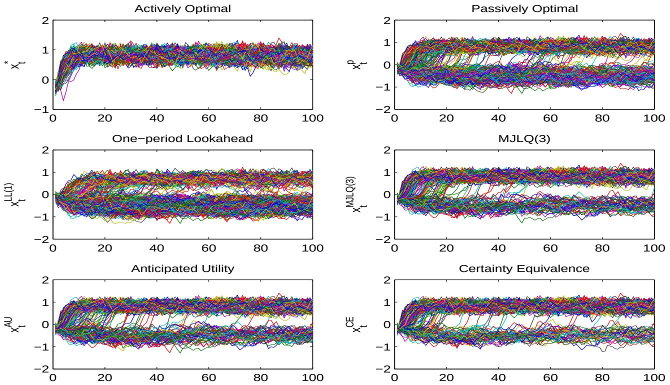


Figure: 13. Simulated multiple time-series of of target state x_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$. Number of simulations: $NMC = 400$.

- Only actively optimal policy does not bifurcate into 2 basins of attraction
- One-period lookahead escapes the two basins most frequently
- Anticipated utility and MJLQ(3) are very similar

Simulated Mean Beliefs

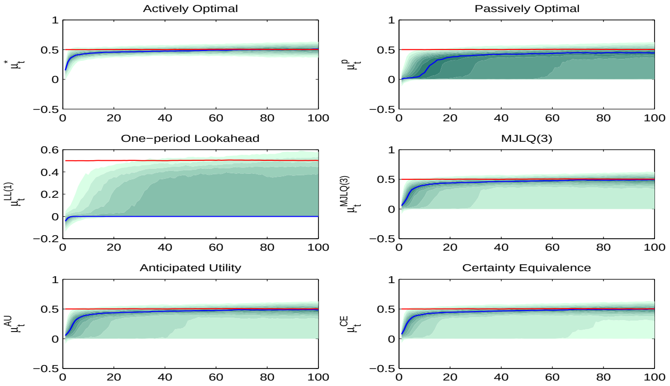


Figure: 14. Evolving distribution of simulated time-series of mean belief μ_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$. Number of simulations: $NMC = 400$.

- Actively optimal policy quickly zooms in into the neighborhood of actual slope realizations
- Actively optimal policy realizations are the tightest band

Simulated Belief Variances

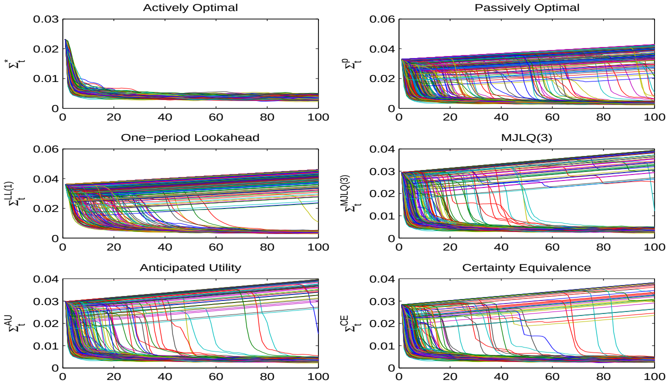


Figure: 15. Simulated multiple time-series of variance of belief Σ_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$. Number of simulations: $NMC = 400$.

- Actively optimal policy is the only one consistently reducing uncertainty
- Learning cannot converge to the truth if $\sigma_\eta^2 > 0$

Simulated Regrets

Regret function

$$C_t = \sum_{\tau=0}^{t} \delta^{\tau} \left((x_\tau - x^*)^2 + \omega (u_\tau - u^*)^2 \right) .$$

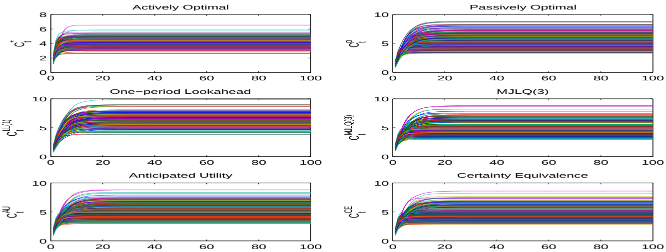


Figure: 16. Simulated multiple time-series of regret function C_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$. Number of simulations: $NMC = 400$.

- Actively optimal policy dominates except in unluckiest cases
- Passively optimal is second best
- Certainty equivalent outcome is the worst

Concluding Remarks

- Six solutions/approximations to Bayesian dual control of drifting coefficient regression
  - certainty equivalent policy
  - anticipated utility control
  - Markov jump linear quadratic approximation (semi-new!)
  - passively optimal policy (new!)
  - one-period limited lookahead
  - actively optimal dual policy
- Common features
  - gradualism
- Distinct features
  - actively optimal policy could be different, especially in the regions of high uncertainty
  - actively optimal policy guards against dismal outcomes
  - active experimentation eliminates escape dynamics dual basins of attraction
  - actively optimal and certainty equivalent policy induce fastest learning
- Anticipated utility and MJLQ(3) are very close, but the former is much easier to compute
- $V^*(S, u^{AU}) - V^{AU}(S, u^{AU})$ could be negative or positive; using $V^{AU}(S, u^{AU})$ could distort inference of the benefit to intentional experimentation

Work in Progress

- New suboptimal policies trading limited lookahead for active prediction of future posterior distributions
  - Multi-step prediction of posterior variance
  - Linearization of filtering distributions in augmented state formulation (EKF, Kendrick (2002))
  - Gaussianization of future posteriors in augmented state formulation (UKF), needed if γ is unknown
- Similar models (up to 6 state variables)
  - $x_{t+1} = \alpha + \beta u_{t+1} + \epsilon_{t+1}$, α, β unknown (Wieland (2000))
  - $x_{t+1} = \alpha + \beta u_{t+1} + \gamma x_t + \epsilon_{t+1}$, β, γ unknown
  - $x_{t+1} = \alpha + \beta_1 u_{t+1} + \beta_2 u_t + \gamma x_t + \epsilon_{t+1}$, β_1, β_2 unknown
  - $x_{t+1} = \alpha + \beta_{t+1} u_{t+1} + \gamma_{t+1} x_t + \epsilon_{t+1}$, β_t, γ_t latent: $\beta_{t+1} = \beta_t + \eta_{t+1}$, $\gamma_{t+1} = \gamma_t + \nu_{t+1}$
  - $x_{t+1} = \alpha + \beta_1 u_{1,t+1} + \beta_2 u_{2,t+1} + \epsilon_{t+1}$, β_1, β_2 unknown
- Additional models
  - Latent volatility models
  - Multivariate target state and cross-equation restrictions
- Numerics and computation
  - GPU accelerated computation
  - Adaptive Smolyak grids
- Theoretical Work
  - Convergence proofs and bounds
  - Expanding uncertainty dimensions
- Applications