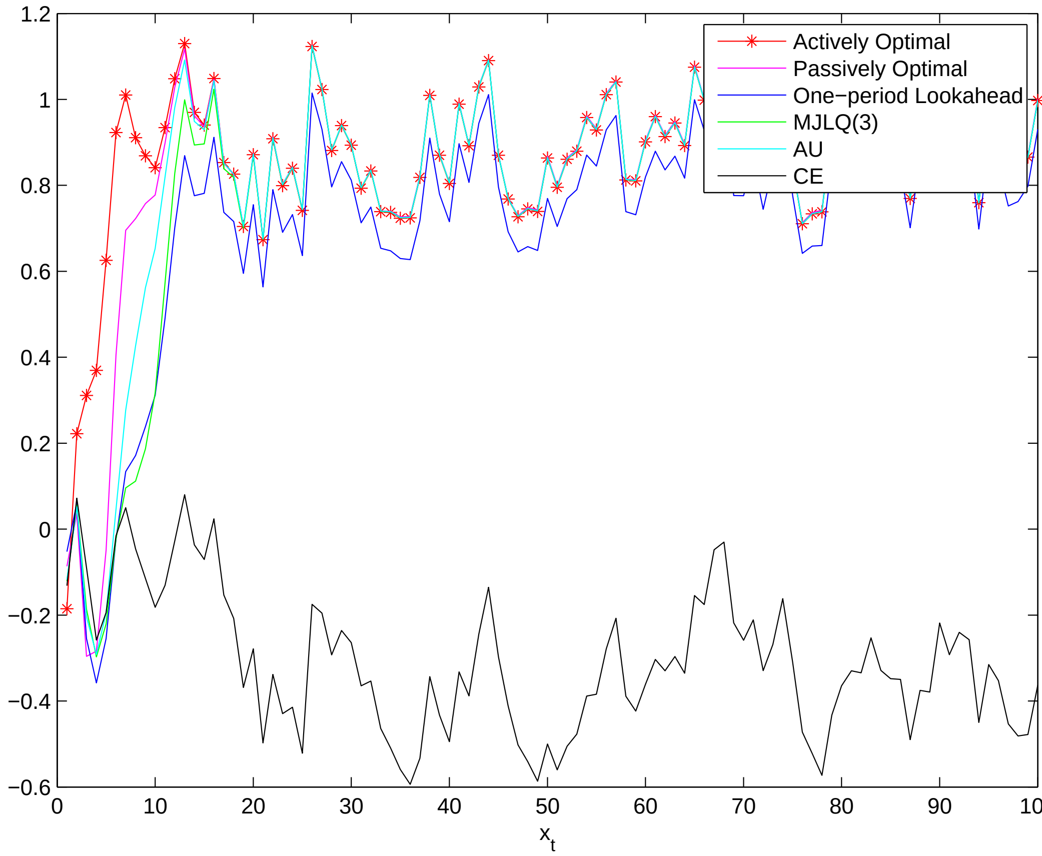

Figure 40: Simulated single time-series of target state x_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$.

the latent coefficient process closer to the truth. Subsequently, the greater knowledge of policy effectiveness is exploited to achieve low regret.

Scatter plots in figures 60 and 61 supplement the evidence for the clear superiority of actively optimal policy by showing that the advantage is uniform – if the shock sequence is fortuitous in that it leads to small cumulative loss under any of the suboptimal policies, the regret under actively optimal policy is even smaller. To reconcile this evidence with only a short distance between expected cost-to-go functions under alternative policies along the state space slices we have shown earlier, we conjecture that the actively optimal policy is robust against really unfortunate but rare draws of stochastic disturbances that may thrust the state into the regions of larger differences among cost-to-go functions and correspondingly larger policy differences. Similar idea is expressed in Amman and Kendrick (1994).

10.7. **Costs of Computation.** One of the important contentions of the approximate dual control methods is their alleged computational superiority over direct numerical dynamic programming. Here we briefly comment on our experiences with the drifting coefficient regression model with this regard.

Table 1 shows CPU time required to evaluate a given policy rule on a grid of a given size. Here we assume that value iteration algorithm for actively optimal and passively optimal

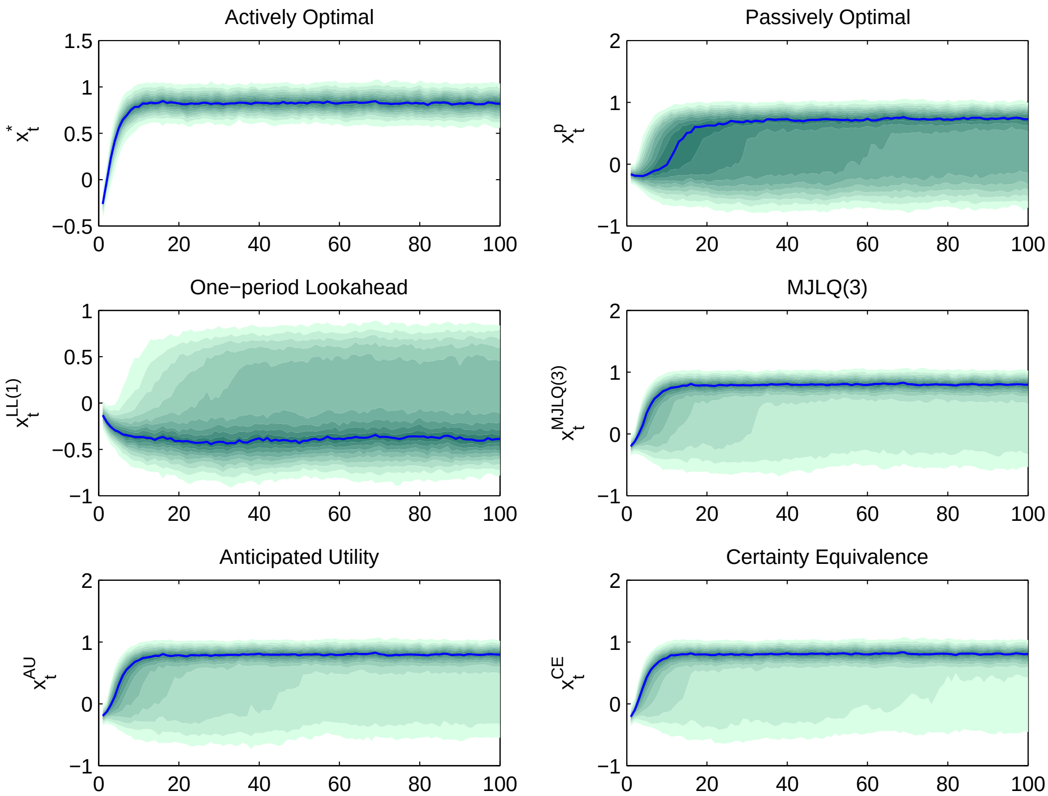

Figure 41: Evolving distribution of simulated time-series of target state x_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$. Number of simulations: $NMC = 400$.

dynamic programs is executed on a $100 \times 100 \times 100$ grid with 100-point univariate Gauss-Hermite quadrature for the former and tensor product 40×40-point Gauss-Hermite quadrature for the latter. The CPU timings were obtained on Intel© Xeon 5160 CPU. Dynamic programming calculations were further accelerated by using parallel dynamic programming using four processors. The timings also account for the fact that certainty equivalent, anticipated utility and MJLQ(3) policies are linear in x_t and hence only tabulating the slope and the intercept is necessary in order to tabulate the policy on the entire three-dimensional grid. In addition, certainty equivalent policy is independent of Σ which reduces the number of evaluations even further. We also assume that quartet of suboptimal policies exploits parallel computation capabilities (four threads) comparable to those available for dynamic programming solutions. For the grids larger than $100 \times 100 \times 100$ we assumed that actively optimal and passively optimal policies are evaluated by linear interpolation. The same is assumed for all other policies for which interpolation is faster than direct evaluation – one-period lookahead and MJLQ(3). The point here is that numerical dynamic programming involves fixed computational costs, while various direct approximations - mostly variable costs. Certainty equivalent control is apparently superior numerically to all the other approaches, yet we have seen that following that type of policy could lead the decision-maker seriously astray. Anticipated utility occupies the middle ground with moderately close approximation which is cheap to compute. Passively optimal policy is the hardest due to two-dimensional integration involved. Among the quartet of approximations studied here MJLQ policy may be at disadvantage – it's very close to AU unless the number of regimes is large, and if it is required to be evaluated at many points in state space (e.g. in the course

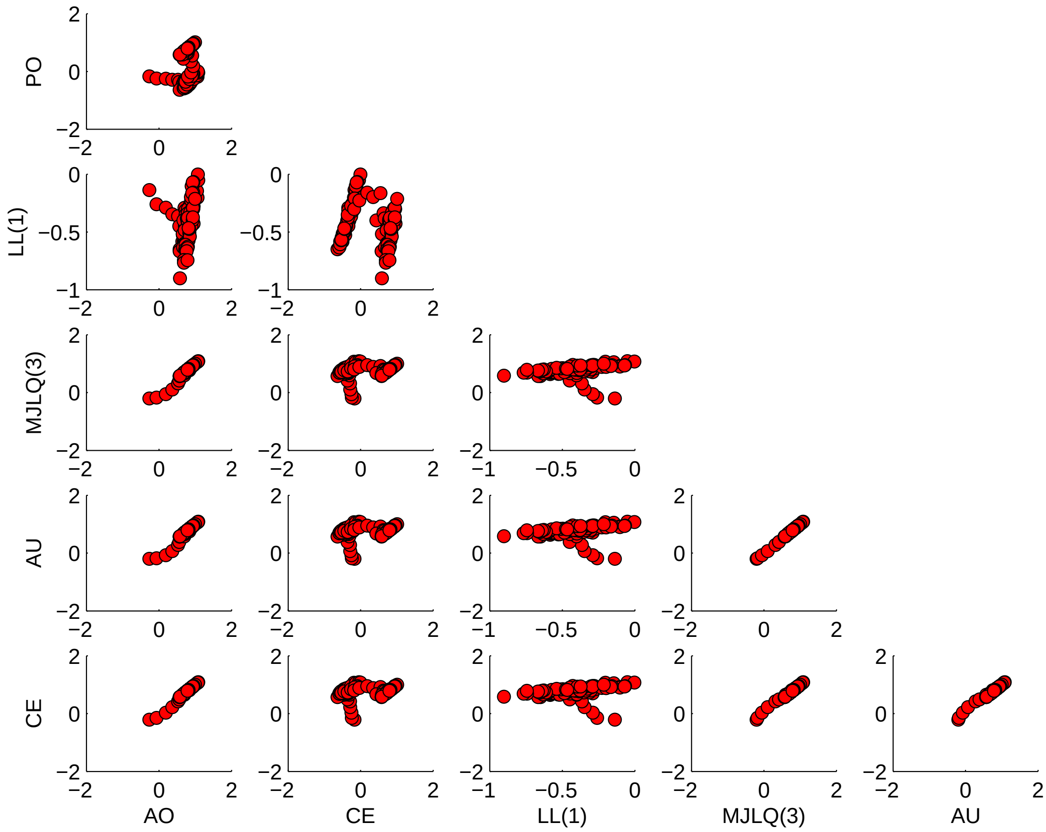

Figure 42: Pairwise comparisons of single simulated time-series of target state x_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$.

of simulation), it may require more time than optimal policy. I conjecture that in problems with more dimensions, especially dimensions of the information state, MJLQ-type control will regain its competitiveness.

An additional aspect to computational cost is its type. Large-scale dynamic programming approaches, especially in the presence of hard-to-quantify discontinuities are not only compute-time bound but also memory limited.

We conclude that while the computational advantage of any given policy is specific to the problem formulation and the desired degree of approximation, the dynamic programming approach need not be summarily discarded.

11. Conclusion

In this paper, we study the problem of the infinite-horizon dual Bayesian control of a regression with drifting sensitivity to the control input, a setup of Beck and Wieland (2002). The focus of the study was two-fold, to expand the set of approximate solutions to include those that take account of drift in the policy effectiveness and to explore the quality of approximation in more detail.

The set of candidate approximations included five policies in addition to the actively optimal policy itself. One approximation, the certainty equivalent policy, ignores parameter uncertainty altogether, including the coefficient drift. It displays the largest amount of accidental experimentation and yet could be seriously lead adrift. The other class of

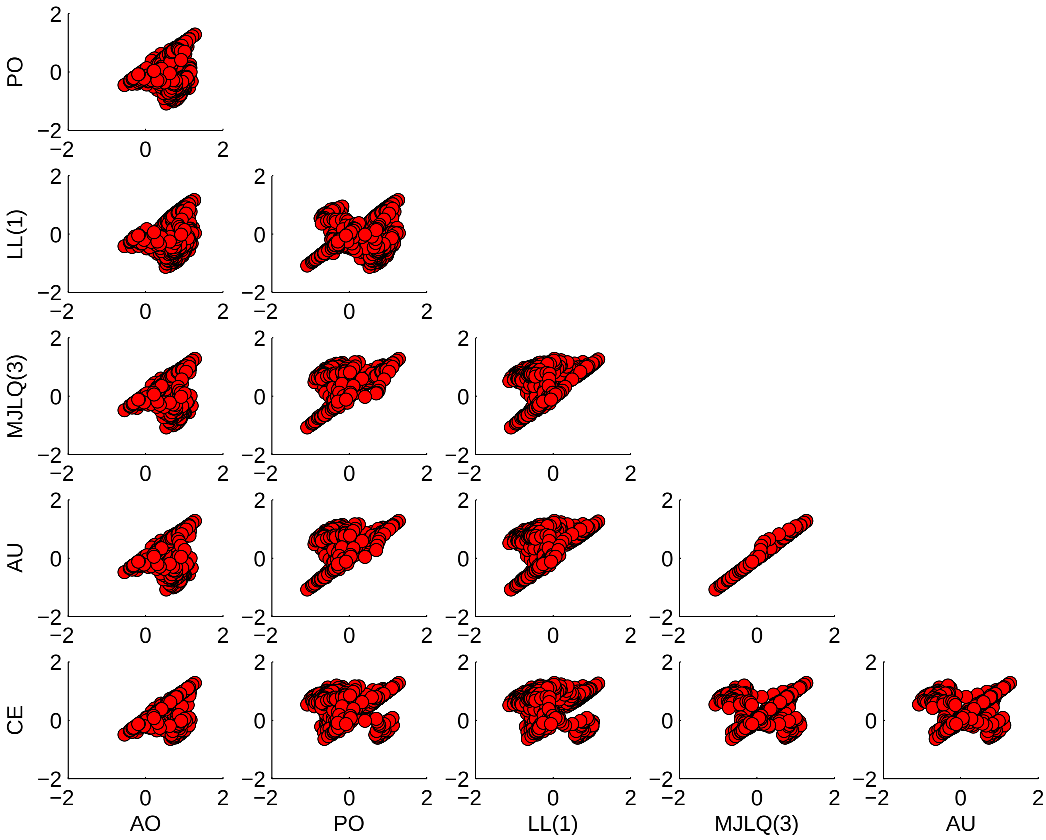

Figure 43: Pairwise comparisons of multiple simulated time-series of target state x_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$. Number of simulations: $NMC = 400$.

	Grid Size		
Policy	1	$100 \times 100 \times 100$	$500 \times 500 \times 500$
Actively Optimal	3.336e+2	3.336e+2	4.878e+3
Passively Optimal	2.596e+3	2.596e+3	7.140e+3
One-period Lookahead	5.667e-3	1.417e+3	5.962e+3
Certainty Equivalent	4.425e-6	1.107e-4	5.531e-4
Anticipated Utility	4.505e-6	1.126e-2	2.816e-1
MJLQ(3)	9.436e-2	2.359e+2	5.897e+3

Table 1: CPU time, T, in seconds, required to evaluate given policy on the grid. Parameter configuration: $\alpha = 0.1$, $\gamma = 0.9$, $\delta = 0.75$, $\omega = 1$, $\sigma_\epsilon^2 = 1.0$, $\sigma_\eta^2 = 0.01$, $x^* = 1$, $u^* = 0$. Grid: non-uniform product grid on $(-2, 2) \times (-3, 3) \times (0.001, 4)$. Dynamic programming algorithms implemented using combination of value and policy iterations.

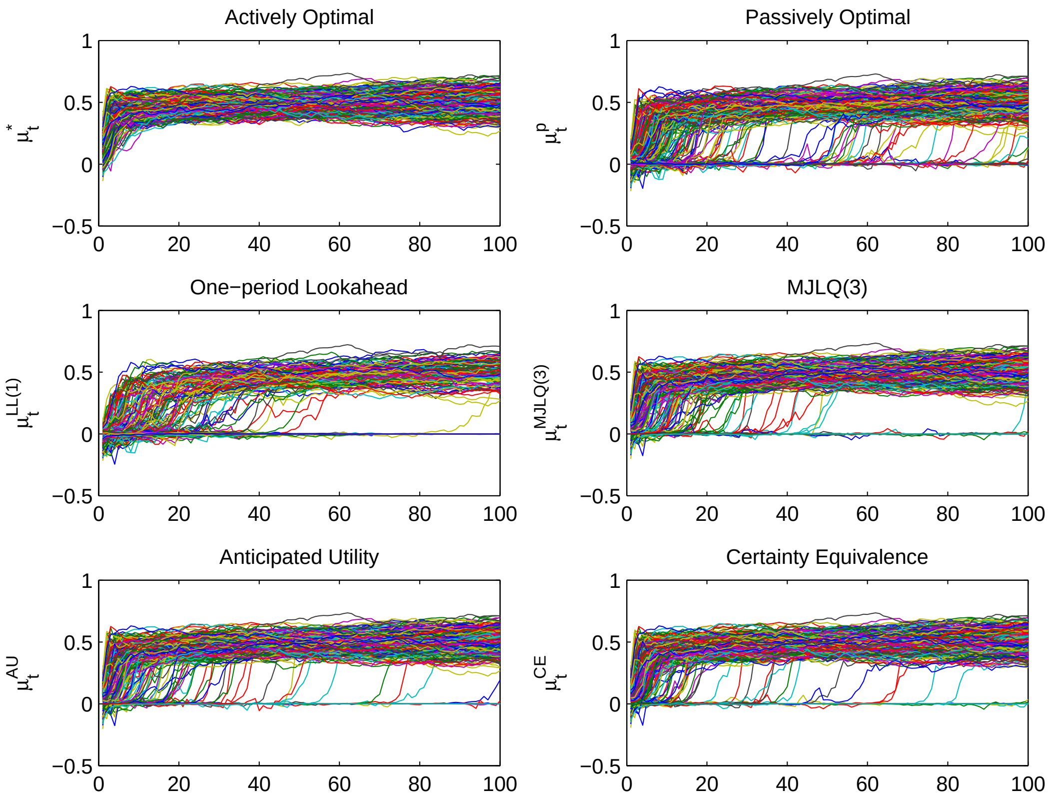

Figure 44: Simulated multiple time-series of mean belief μ_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$. Number of simulations: $NMC = 400$.

approximations is represented by the one-period lookahead. This is an actively adaptive policy that acknowledges both the parameter uncertainty and the coefficient drift but is semi-myopic in that it only solves two-period dynamic program, ignoring the costs accruing further into the future. Even with relatively low discount factor, one-period lookahead policy could induce outcomes that do not resemble those of optimal policy and which may not be shared by other policies. For example, simulated one-period lookahead controls are attracted by the no-intervention policy ($u_t = 0$) nearly 50% of the time, while at the same time the bifurcation of the state paths into two branches is least pronounced. The third class of approximations is comprised of the three policies that take parameter uncertainty into consideration but eschew the active experimentation motive. They differ in how seriously they treat the coefficient drift. Anticipated utility approach ignores the drift, MJLQ policy treats coefficient drift as if it follows finite state Markov chain, while passively optimal policy recognizes that the support of the drift is continuous.

Our calculations both detect common features and discern contrasting elements among the six policy rules. On the one hand, casual inspection of the optimal cost-to-go function and optimal Q-factors of the other five policies indicate that only the certainty equivalent policy rule incurs significantly larger expected cost, with the expected loss of the four remaining suboptimal policies being close to that under the actively optimal policy, at least in the "central" regions in the state space. On the other hand, the actively optimal policy consistently outperforms all of the alternatives by a noticeable margin. The simulation advantage can be traced back to the remarkable bifurcation of the simulated state dynamics into two distinct branches under all suboptimal policies but not under actively optimal

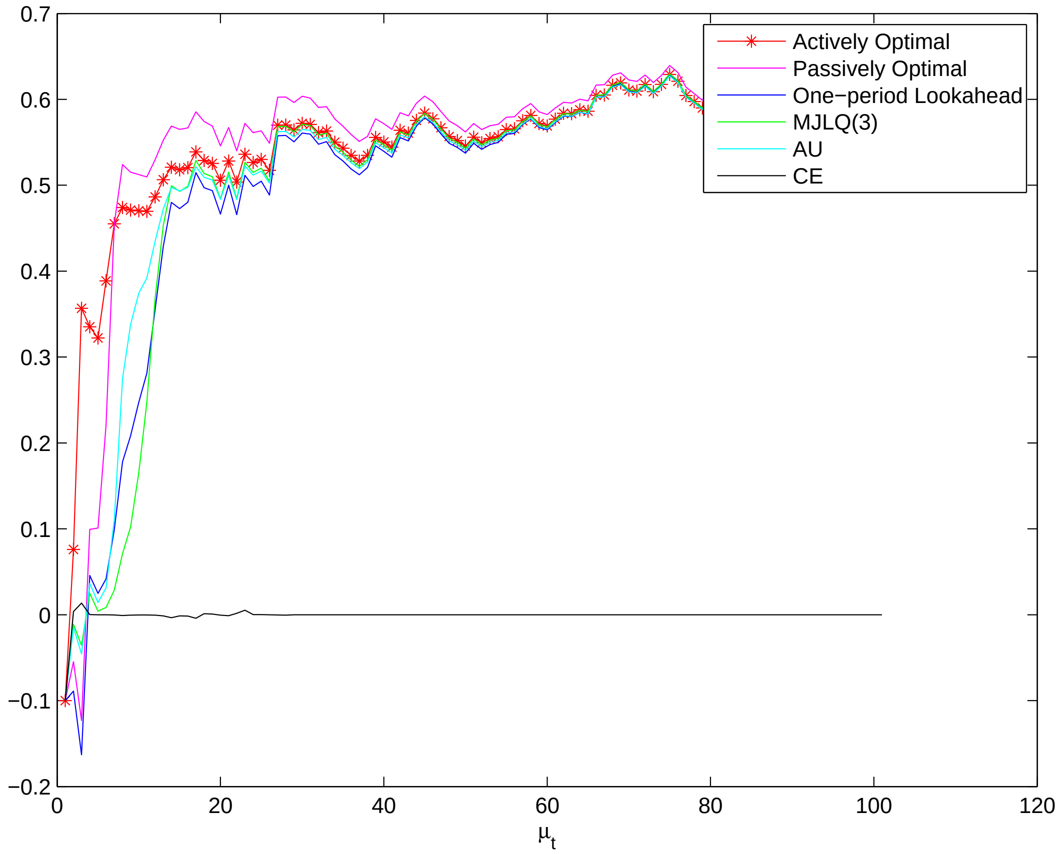

Figure 45: Simulated single time-series of mean belief μ_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$.

solution. Five suboptimal approximations share a kind of escape-like dynamics between the two basins of attraction, not unlike those observed by Hughes and Jacobs (1974) and systematically studied in Cho, Williams, and Sargent (2002); Sargent and Williams (2005) among others. As a result, mean beliefs under the actively optimal policy move rapidly towards the unobserved true values for practically all simulated paths, while under alternatives the probability of traipsing far away from true values remains noteworthy. Potential lack of tightening of the posterior belief (or even gradual widening) is then possible. Because the one-period lookahead policy displays the smallest amount of branching and is the only active but suboptimal policy, we can possibly attribute elimination of two distinct basins of attraction to the active experimentation. Our results also indicate that MJLQ control is a modest improvement over anticipated utility policy in terms of reaching the optimal performance index in the class of passively adaptive policies. In simulation, they are very hard to distinguish, along practically every dimension of interest, at least for the moderate number of regimes, while anticipated utility is much easier to compute. Both anticipated utility and MJLQ controls are not very different from the optimal passively adaptive policy. Moreover, the benefit to the active experimentation over most passively adaptive policies is relatively small. On the other hand, more sophisticated learning techniques and probing seem important for reducing the likelihood of large mistakes and escape-like dynamics. Only the actively optimal policy appears robust to unlikely unlucky draws of stochastic disturbances.

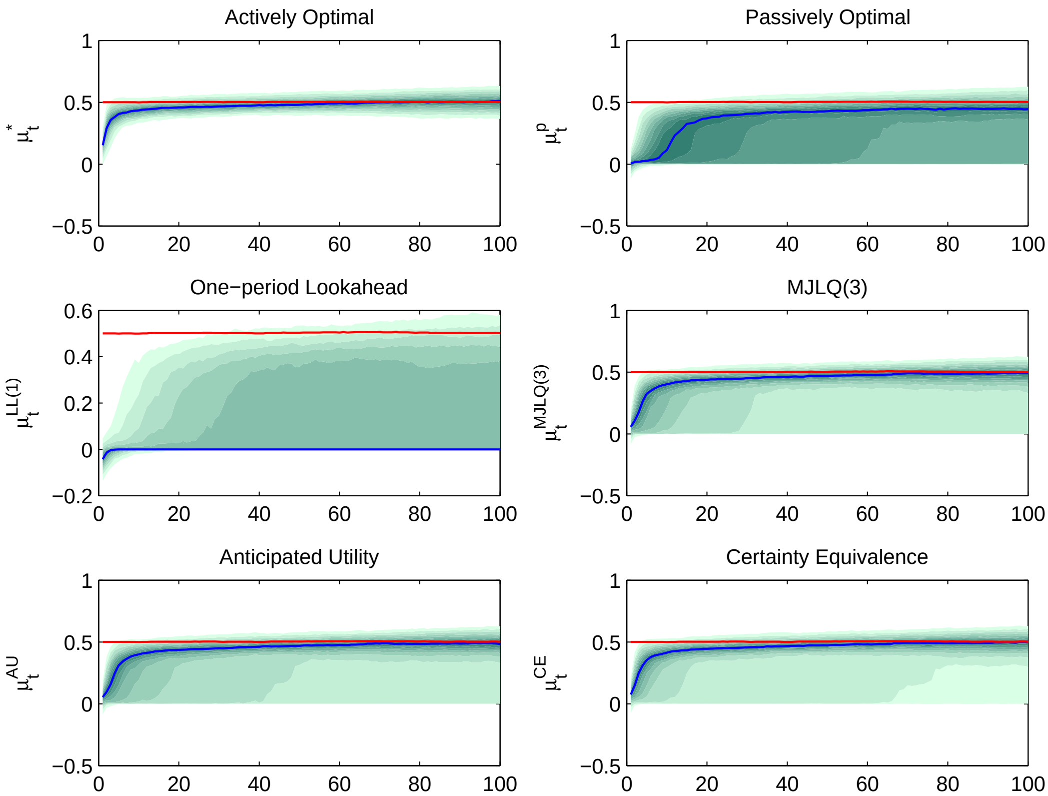

Figure 46: Evolving distribution of simulated time-series of mean belief μ_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$. Number of simulations: $NMC = 400$.

Along the way, we found that there's no definitive relationship between optimal Q-factor of anticipated utility policy and the cost-to-go function optimized by the anticipated utility control. Comparison of the latter with the optimal cost-to-go function can therefore distort the evaluation of the benefit accruing to the active experimentation. In light of this finding, we suggest that results such as Cogley, Colacito, and Sargent (2007) on the relative insignificance of the intentional experimentation component of policy be interpreted with caution.

In addition to being of interest in their own right as reasonably rational decision rules and as relative to the optimal dual policy, suboptimal policies serves as good starting points for more computationally demanding actively adaptive solutions. Indeed, using the value of the myopic policy instead of the zero continuation value eliminates one iteration from the standard value iteration dynamic programming algorithm.

Whereas our preliminary findings are not comprehensive, they suggest that the role of probing and taking account of future dynamic parameter uncertainty is not so much in the reduction of expected stream of losses, but guarding against truly dismal outcomes. If my results remain valid in more realistic and empirically relevant settings, the design of optimal economic policies in general and monetary policy in particular may shift its emphasis. Analysis of the benefits to the experimentation and adaptive learning in broader settings is a topic of on-going research.

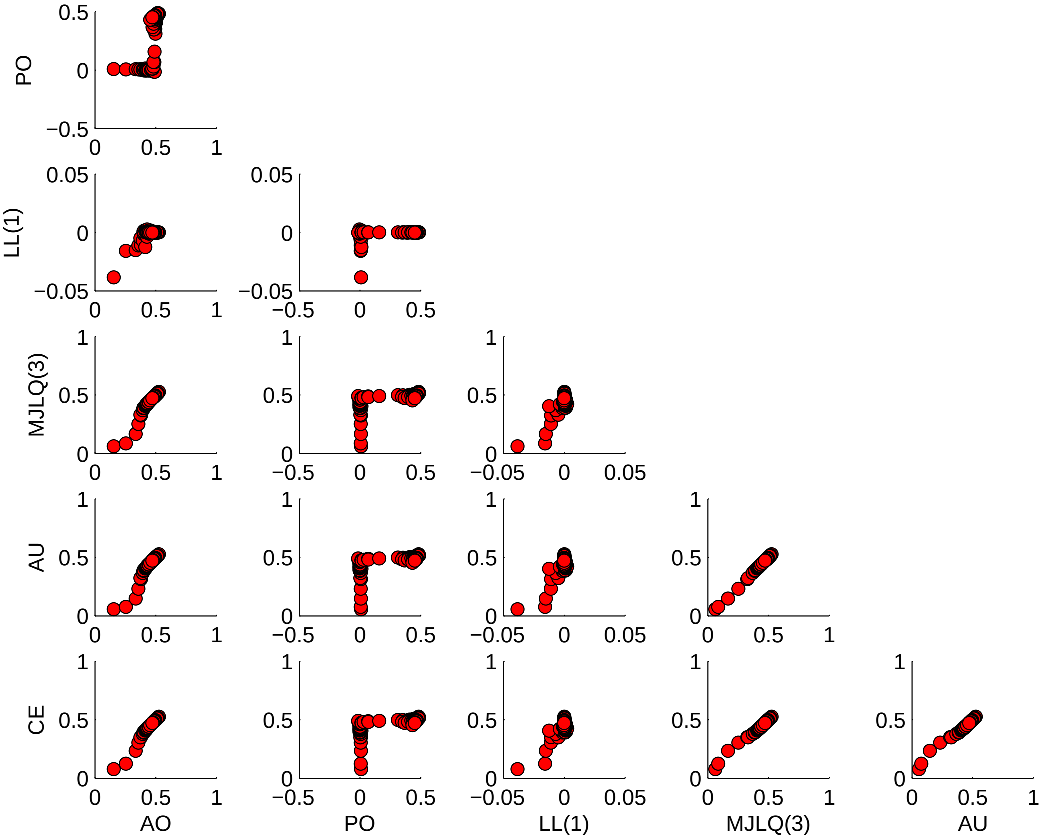

Figure 47: Pairwise comparisons of single simulated time-series of mean belief μ_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$.

Appendix A. Details of Computation

Actively optimal and passively optimal policies and cost-to-go functions are represented by means of linear interpolation on the non-uniform tensor product grid in the state space. The non-uniform grid is designed to place grid-points more densely in the areas of high curvature, namely in the vicinity of $x = x^*$ and $\mu = 0$. The grid is uniform along Σ dimension. Although, in principle, the state space is unbounded, we restricted our attention to the three-dimensional cube. The boundaries were chosen via experimentation to ensure that high curvature regions are completely covered and that all simulated sequences originating sufficiently deep inside the cube remain there for the entire time span of a simulation. Once the boundaries were chosen we used 200 points along each dimension for the actively optimal dynamic program and $100 \times 64 \times 64$ for the passively optimal dynamic program.[14]

[14]The tensor product grid was necessitated by known potential for non-convexities of the cost-to-go function with respect to the policy variable, which imply non-smooth shape of the optimal cost-to-function with respect to the state variables and, hence, discontinuities of the policy function. These difficulties precludes use of more efficient ways to combat the curse of dimensionality such as Smolyak sparse grid algorithm (Krueger and Kubler, 2004) or projection methods (Judd, 1998) as these rely heavily on function continuity. Adaptive Smolyak sparse grid methods that offer heuristics designed to concentrate most of the points in the directions that have the steepest gradient or have discontinuities have been proposed recently (Gerstner and Griebel, 1998; Novak, Ritter, Schmitt, and Steinbauer, 1999; Hegland, 2001; Bungartz and Dirnsorfer, 2003; Klimke and Wohlmuth, 2005). Application of these methods to imperfect information dynamic programs of higher dimension is a subject of future research.

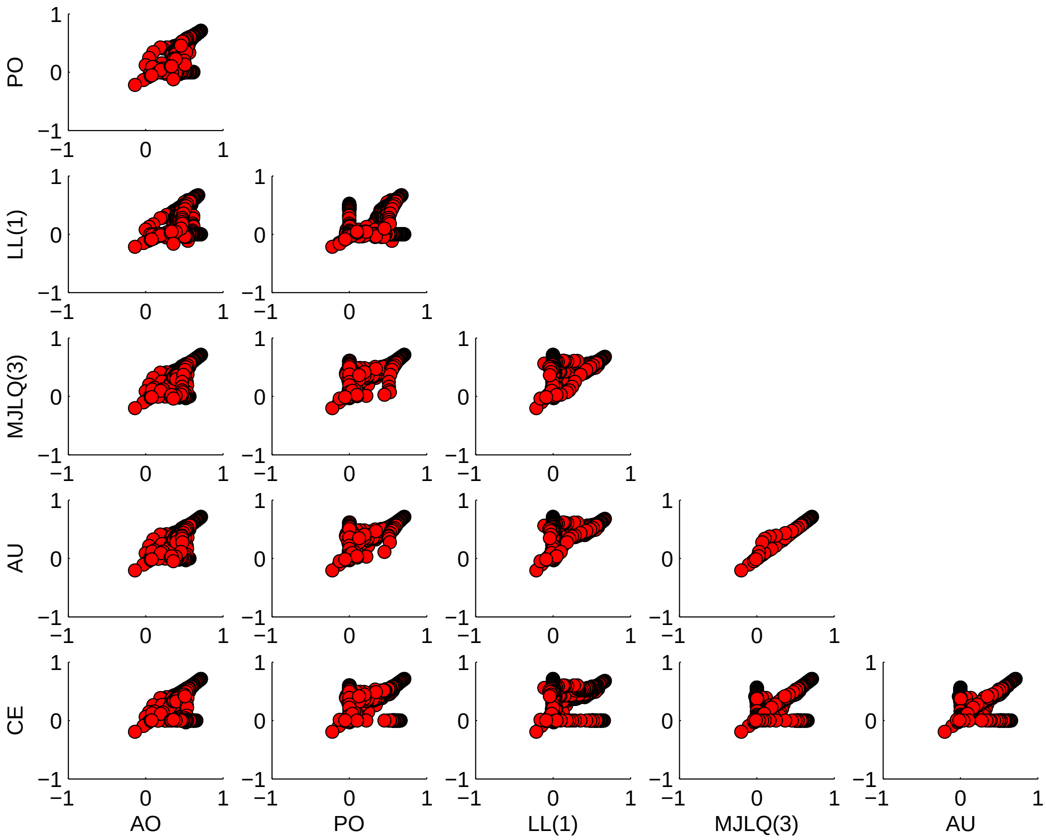

Figure 48: Pairwise comparisons of multiple simulated time-series of mean belief μ_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$. Number of simulations: $NMC = 400$.

Notice also that the optimal value function in the dynamic program with passive learning is non-stationary with respect to the variance of beliefs which drifts to infinity. Receding variance boundary (by analogy with receding horizon control) was used until the solution was no dependent on the variance boundary within $10e - 3$ relative error.

The computation of actively optimal policy is done via the combination of value and policy iterations. Unlike the standard policy iteration where a single policy improvement step (single value iteration) is alternated with a single policy evaluation step (by performing policy iterations to convergence), we use multiple policy improvement steps to speed up convergence. Preliminary experimentation has indicated that it is best in terms of computing time to use four policy improvement steps per one policy evaluation when $\delta = 0.75$. To further accelerate convergence we tried the midpoint of McQueen-Porteus error bounds (Bertsekas, 2001; Rust, 1996) as a better next guess for the value function. The acceleration was not successful and consequently was abandoned. During the minimization step of the actively optimal and passively optimal dynamic programs we used `Fortran` subroutine `VD04AD` from HSL Archive (formerly Harwell Subroutine Library), see HSL (2002). `VD04AD`, originally due to M.J.D. Powell, finds a minimum of a smooth function of a single variable using safeguarded quadratic interpolation with no attempt being made to avoid local minima which may not be global minimum. Due to the local nature of optimization using good starting values is important to locate the correct minimum and to accelerate performance. Once we ascertained that anticipated utility policy closely mimic the two optimal policies,

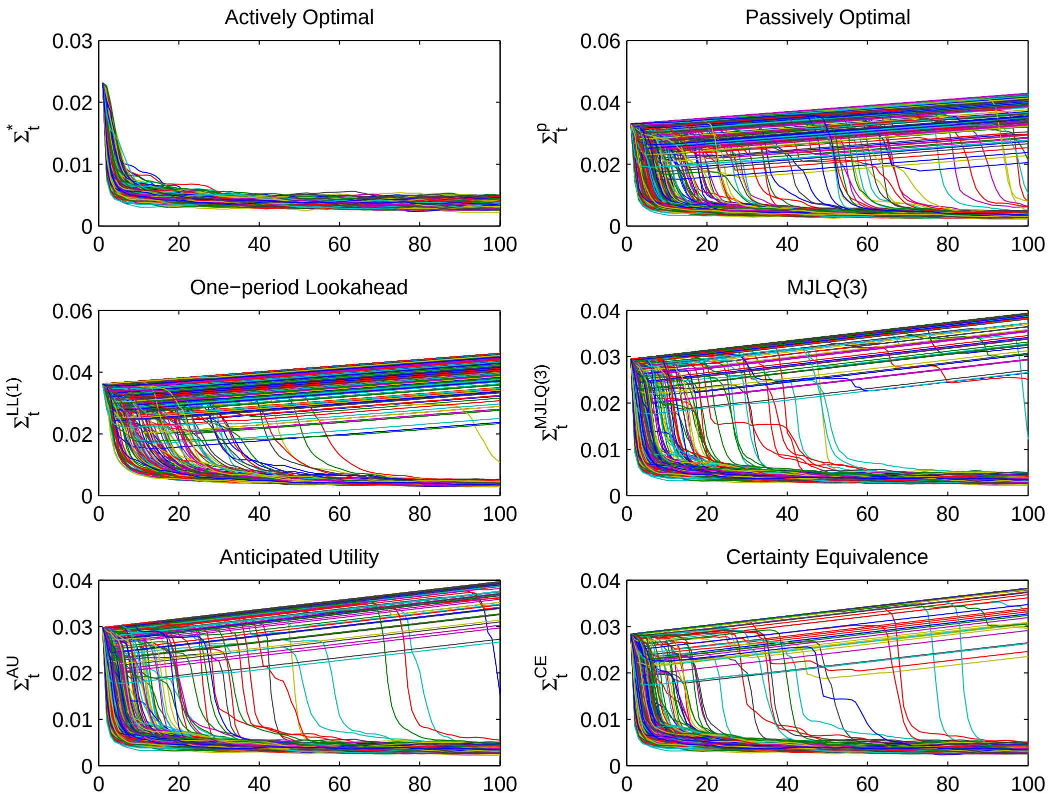

Figure 49: Simulated multiple time-series of variance of belief Σ_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$. Number of simulations: $NMC = 400$.

we used these as intelligent starting points for the minimization procedure on the right hand side of Bellman equation. In addition, we used "hot start" whenever possible. Thus, exploration of parametric dependence of the optimal policies was sped up by using as a starting guess the value of the policy corresponding to the nearby parameter value.

Gauss-Hermite quadrature with 100 nodes was used to evaluate the expectation of the value function on the right hand side of the Bellman equation. For the passively optimal problem that involves double integration we used tensor product integration with 25 nodes along each dimension. Obviously, the integration methods specifically geared towards double integration against bivariate Gaussian density would be more efficient numerically.

Finally, we parallelized iterative sweeps of the state-space grid during the value and policy iterations for use with multi-threaded computer hardware. This was done with `OpenMP` directives for `Fortran` (Chandra, Menon, Dagum, and Kohr, 2000). We found that the performance scaling with the number of parallel processors (up to four CPUs) was fairly good. The code for the two dynamic program was implemented in `Fortran-90`.

Calculation of the optimal Q-factors of the various alternative policies used similar dynamic programming infrastructure in terms of functional approximation, parallelization and numerical integration but only using the policy iteration steps.

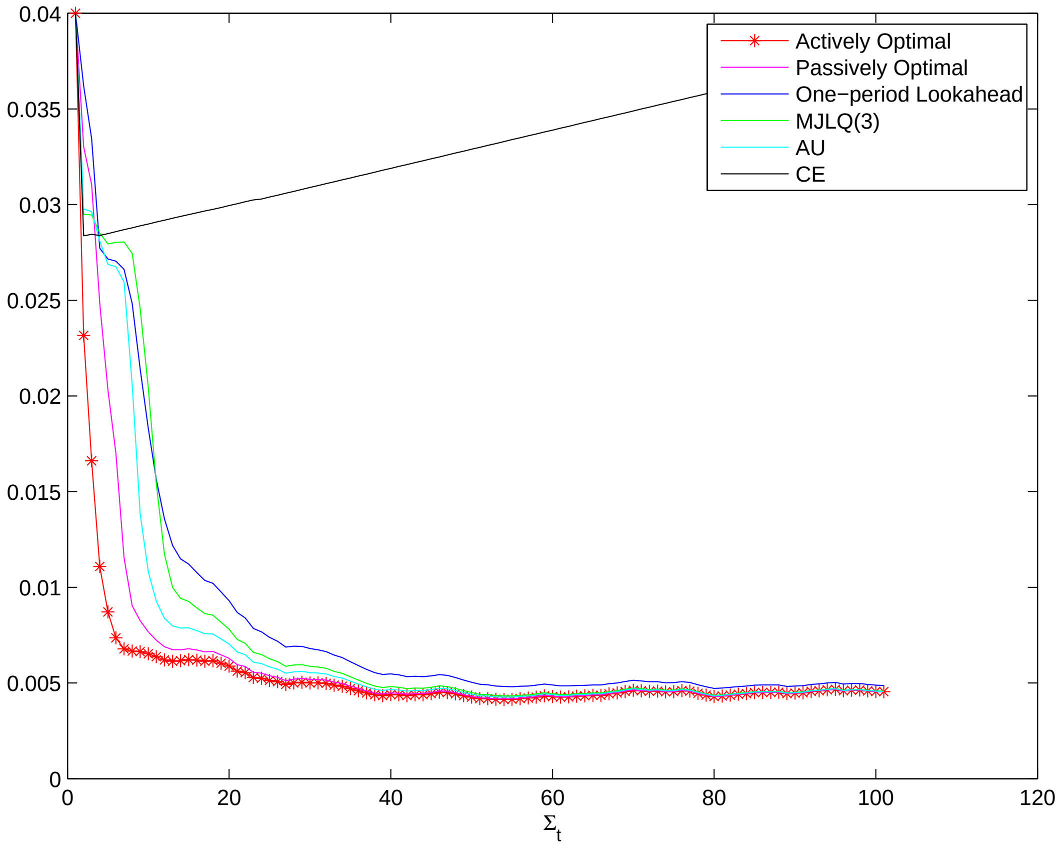

Figure 50: Simulated single time-series of variance of belief Σ_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$.

Other policies were implemented directly, in `Matlab`. Robust univariate minimizer `fminsearch` was used for the limited lookahead calculation.[15] The algorithm for MJLQ was implemented as described in the main body of the article.

References

Amman, H. M., and D. A. Kendrick (1994): "Active Learning: Monte Carlo Results," *Journal of Economic Dynamics and Control*, 18(1), 119–124.

Beck, G. W., and V. Wieland (2002): "Learning and control in a changing economic environment," *Journal of Economic Dynamics and Control*, 26(9-10), 1359–1377.

Bertsekas, D. P. (2001): *Dynamic Programming and Optimal Control*, vol. 2. Athena Scientific, Nashua, NH, 2 edn.

——— (2005): *Dynamic Programming and Optimal Control*, vol. 1. Athena Scientific, Nashua, NH, 3 edn.

Brock, W. A., S. N. Durlauf, and K. D. West (2007): "Model uncertainty and policy evaluation: Some theory and empirics," *Journal of Econometrics*, 136(2), 629–664.

Bungartz, H.-J., and S. Dirnsorfer (2003): "Multivariate quadrature on adaptive sparse grids," *Computation*, 71(1), 89–114.

[15]We also experimented with alternative global optimization approaches such as direct search and simulated annealing. The results were essentially identical but required more computing time.

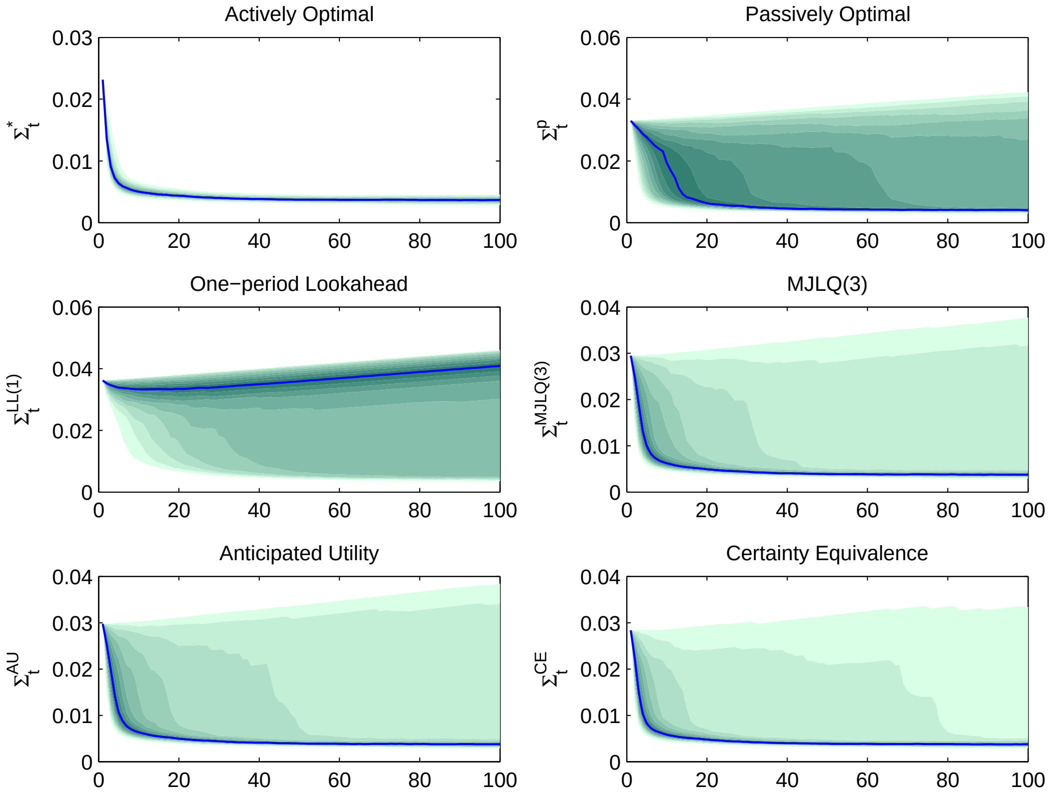

Figure 51: Evolving distribution of simulated time-series of variance of belief Σ_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$. Number of simulations: $NMC = 400$.

Canova, F. (2006): "Monetary Policy and the Evolution of the US Economy," unpublished manuscript, London Business School.

——— (2007): *Methods for Applied Macroeconomic Research*. Princeton University Press, Princeton and Oxford.

Chandra, R., R. Menon, L. Dagum, and D. Kohr (2000): *Parallel Programming in OpenMP*. Morgan Kaufmann.

Cho, I.-K., N. Williams, and T. J. Sargent (2002): "Escaping Nash Inflation," *The Review of Economic Studies*, 69(1), 1–41.

Cogley, T., R. Colacito, and T. J. Sargent (2007): "Benefits from U.S. Monetary Policy Experimentation in the Days of Samuelson and Solow and Lucas," *Journal of Money, Credit and Banking*, 39(S1), 67–99.

Cogley, T., S. Morozov, and T. J. Sargent (2005): "Bayesian fan charts for the UK inflation: Forecasting and sources of uncertainty in an evolving monetary system," *Journal of Economic Dynamics and Control*, 29(11), 1893–1925.

Cogley, T., and T. J. Sargent (2001): "Evolving Post-World War II U.S. Inflation Dynamics," in *NBER Macroeconomic Annual*, ed. by B. Bernanke, and K. S. Rogoff, vol. 16, pp. 331–372. MIT Press.

——— (2005): "The conquest of US inflation: Learning and robustness to model uncertainty," *Review of Economic Dynamics*, 8(2), 528–563.

Costa, O. L. V., M. D. Fragoso, and R. P. Marques (2005): *Discrete-Time Markov Jump Linear Systems*, Probability and its Applications. Springer-Verlag, London, 1 edn.

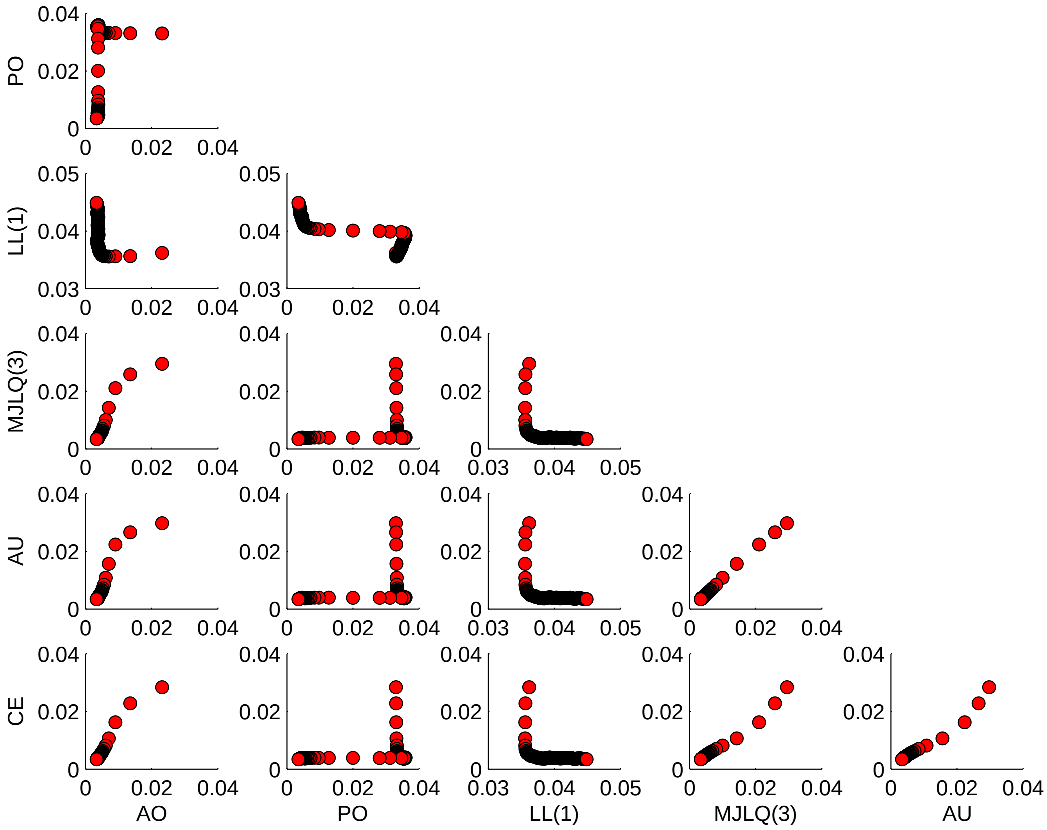

Figure 52: Pairwise comparisons of single simulated time-series of variance of belief Σ_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$.

do Val, J. B. R., J. C. Geromel, and O. L. V. Costa (1998): "Uncoupled Riccati iterations for the linear quadratic control problem of discrete-time Markov jump linear systems," *IEEE Transactions on Automatic Control*, 43(12), 1727–1733.

Easley, D., and N. M. Kiefer (1988): "Controlling a Stochastic Process with Unknown Parameters," *Econometrica*, 56(5), 1045–1064.

Feldbaum, A. A. (1960a): "Theory of dual control (I)," *Automation and Remote Control*, 21(9), 1240–1249.

——— (1960b): "Theory of dual control (II)," *Automation and Remote Control*, 21(11), 1453–1464.

——— (1961a): "Theory of dual control (III)," *Automation and Remote Control*, 22(1), 3–16.

——— (1961b): "Theory of dual control (IV)," *Automation and Remote Control*, 22(2), 129–143.

Gerstner, T., and M. Griebel (1998): "Numerical integration using sparse grids," *Numerical Algorithms*, 18(3-4), 209–232.

Gourieroux, C., and J. Jasiak (1999): "Nonlinear Persistence and Copersistence," Working paper, CREST and CEPREMAP.

Hamilton, J. D. (1994): *Time Series Analysis*. Princeton University Press.

Hansen, L. P., and T. J. Sargent (2004): "Recursive Models of Dynamic Linear Economies," Unpublished Manuscript.

Hegland, M. (2001): "Adaptive sparse grids," in *Proceedings of CTAC*, Brisbane.

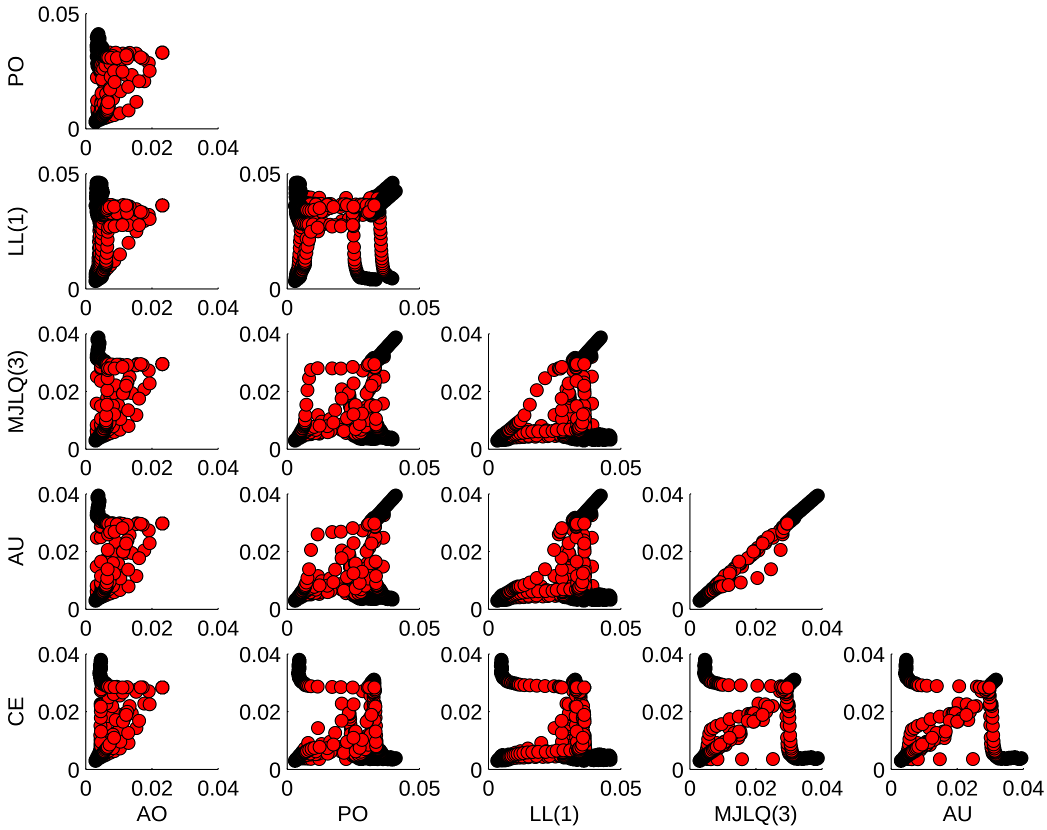

Figure 53: Pairwise comparisons of multiple simulated time-series of variance belief Σ_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$. Number of simulations: $NMC = 400$.

Hinderer, K. (1970): *Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter*, vol. 33 of *Notes in Operations Research and Mathematical Systems*. Springer.

HSL (2002): "A collection of Fortran codes for large scale scientific computation," Technical report, Numerical Analysis Group, Rutherford Appleton Laboratory, http://www.numerical.rl.ac.uk/hsl.

Hughes, D. J., and O. L. R. Jacobs (1974): "Turn-off, Escape and Probing in Nonlinear Stochastic Control," in *Proceedings of IFAC Symposium on Stochastic Control*, pp. 343–352, Budapest, Hungary.

Judd, K. (1998): *Numerical Methods in Economics*. MIT Press.

Kendrick, D. A. (2005): "Stochastic control for economic models: past, present and the paths ahead," *Journal of Economic Dynamics and Control*, 29(1-2), 3–30.

Klimke, A., and B. Wohlmuth (2005): "Algorithm 847: spinterp: Piecewise Multilinear Hierarchical Sparse Grid Interpolation in MATLAB," *ACM Transactions on Mathematical Software*, 31(4).

Krueger, D., and F. Kubler (2004): "Computing equilibrium in OLG models with stochastic production," *Journal of Economic Dynamics and Control*, 28(7), 1411–1436.

Kumar, P. R. (1985): "A Survey of Some Results in Stochastic Adaptive Control," *SIAM Journal on Control and Optimization*, 23(3), 329–380.

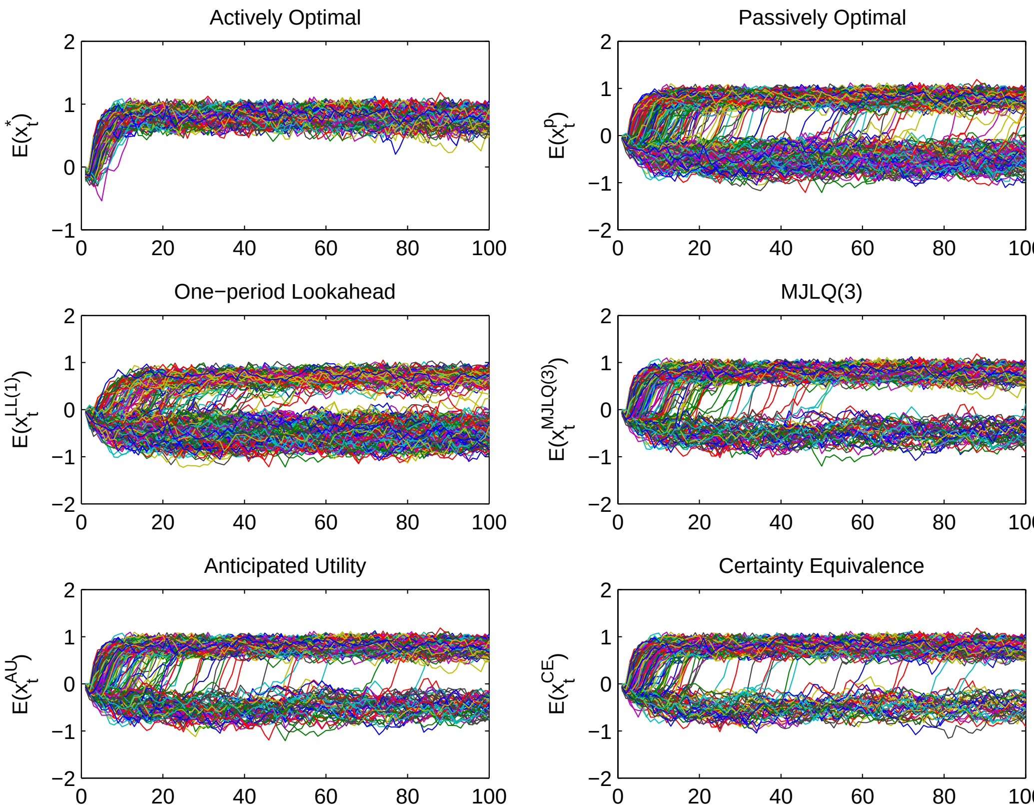

Figure 54: Simulated multiple time-series of expected target state $\mathbb{E}_t(x_t)$ under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$. Number of simulations: $NMC = 400$.

Novak, E., K. Ritter, R. Schmitt, and A. Steinbauer (1999): "On an interpolatory method for high dimensional integration," *Journal of Computational and Applied Mathematics*, 112(1-2), 215–228.

Rust, J. (1996): "Numerical dynamic programming in economics," in *Handbook of Computational Economics*, ed. by H. M. Amman, D. A. Kendrick, and J. Rust, vol. 1 of *Handbooks in Economics*, chap. 14, pp. 619–730. Elsevier North-Holland, Amsterdam, 1 edn.

Sargent, T. J., and N. Williams (2005): "Impacts of Priors on Convergence and Escape from Nash Inflation," *Review of Economic Dynamics*, 8(2), 360–391.

Sims, C. A., and T. Zha (1999): "Error Bands for Impulse Responses," *Econometrica*, 67(5), 1113–1156.

Svensson, L. E. O., and N. Willams (2007): "Bayesian and Adaptive Optimal Policy under Model Uncertainty," Working Paper, Princeton University.

Svensson, L. E. O., and N. Williams (2005): "Monetary Policy with Model Uncertainty: Distribution Forecast Targeting," Working Paper, Princeton University.

Morgan Stanley
E-mail address: Sergei.Morozov@morganstanley.com

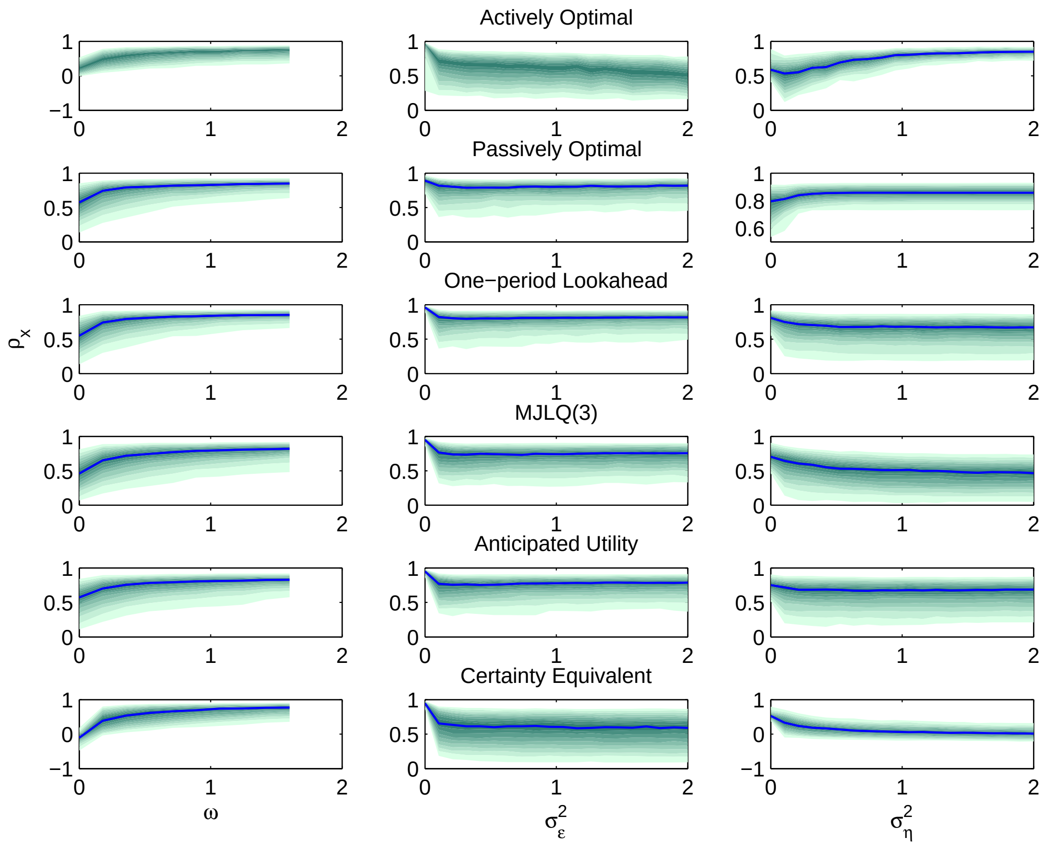

Figure 55: Dependence of state persistence ρ_x on parameters ω, σ^2_ϵ and σ^2_η under different policies. Parameter values (unless varying): $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma^2_\epsilon = 1.0$, $\sigma^2_\eta = 0.01$, $\omega = 0.5$, $x^* = 1.0$, $u^* = 0$. Starting value: $x_0 = 0$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$. Number of simulations: $NMC = 400$.

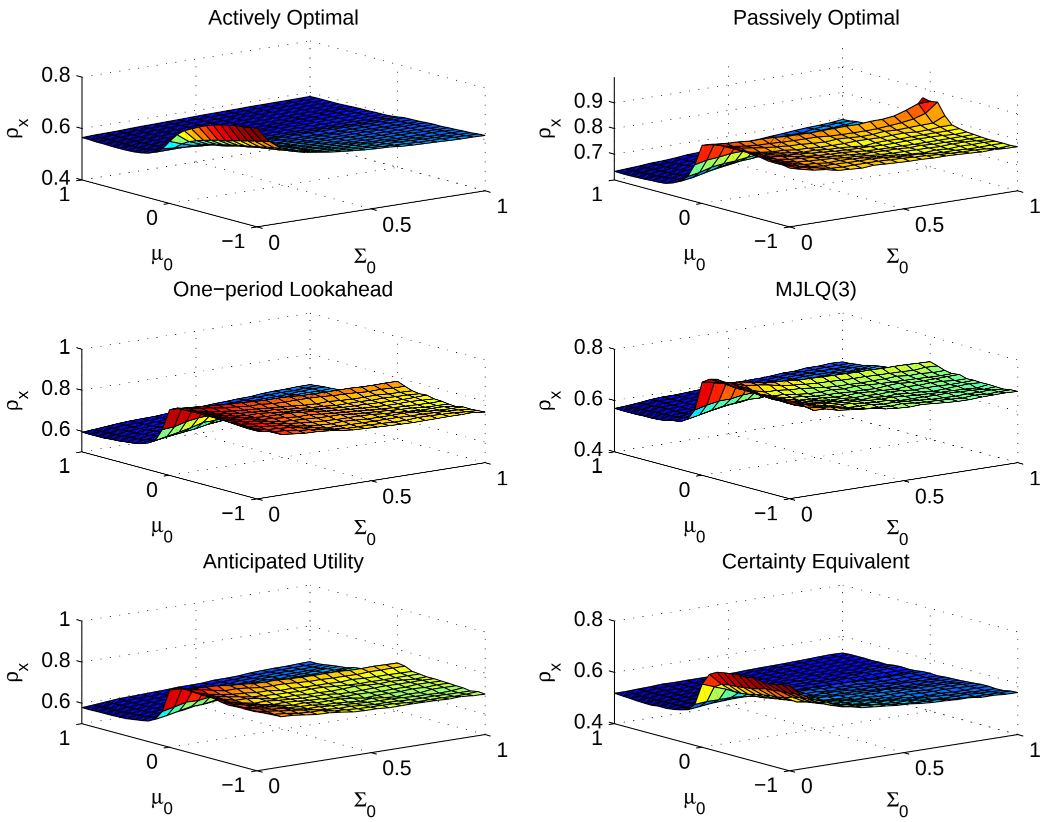

Figure 56: Dependence of state persistence ρ_x on prior beliefs under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 1.0$, $\sigma_\eta^2 = 0.01$, $\omega = 0.5$, $x^* = 1.0$, $u^* = 0$. Starting value: $x_0 = 0$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$. Number of simulations: $NMC = 400$.

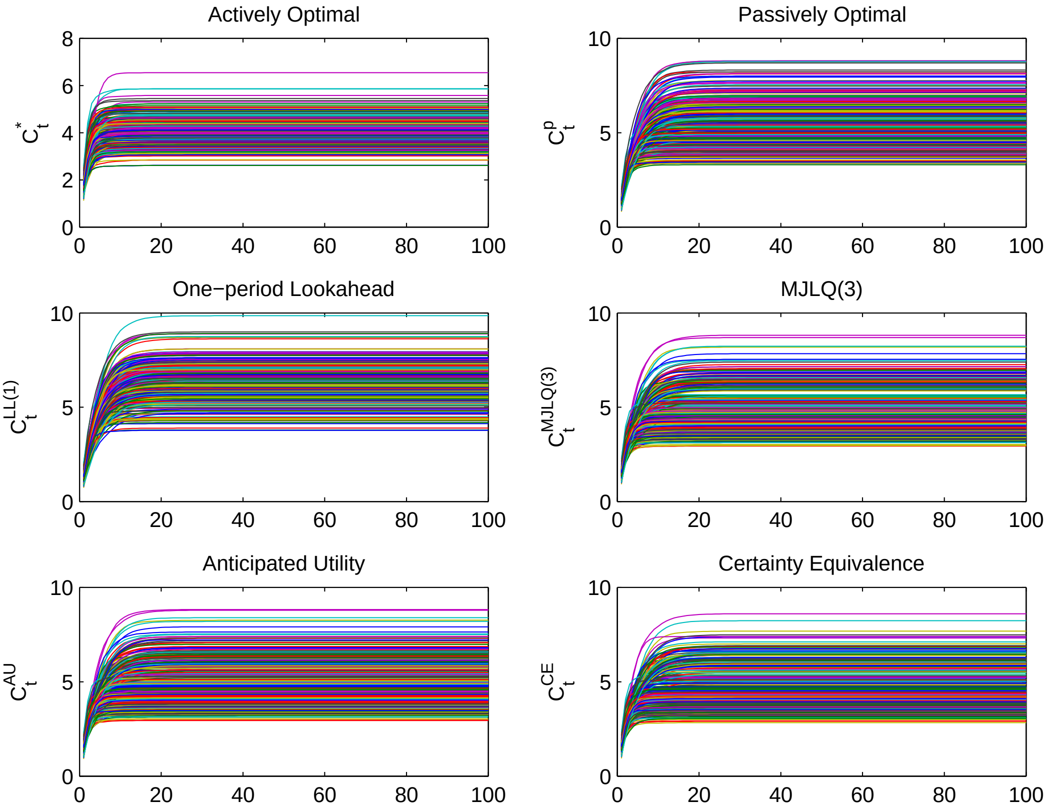

Figure 57: Simulated multiple time-series of regret function C_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$. Number of simulations: $NMC = 400$.

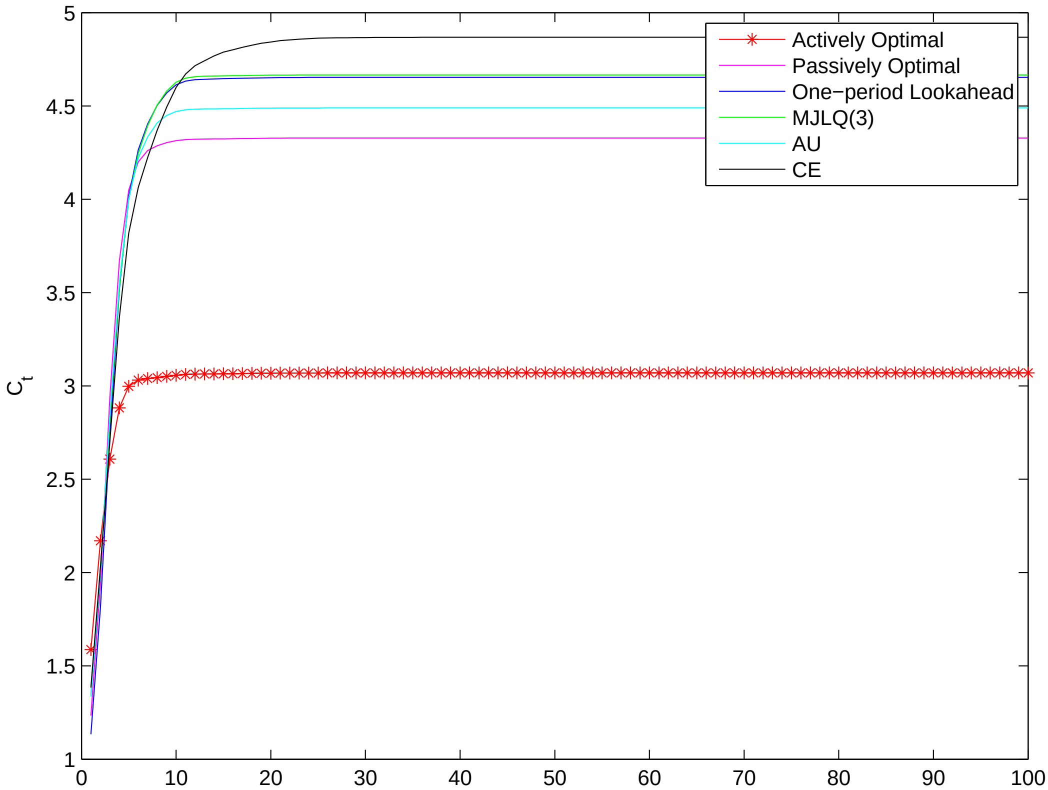

Figure 58: Simulated single time-series of regret function C_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$.

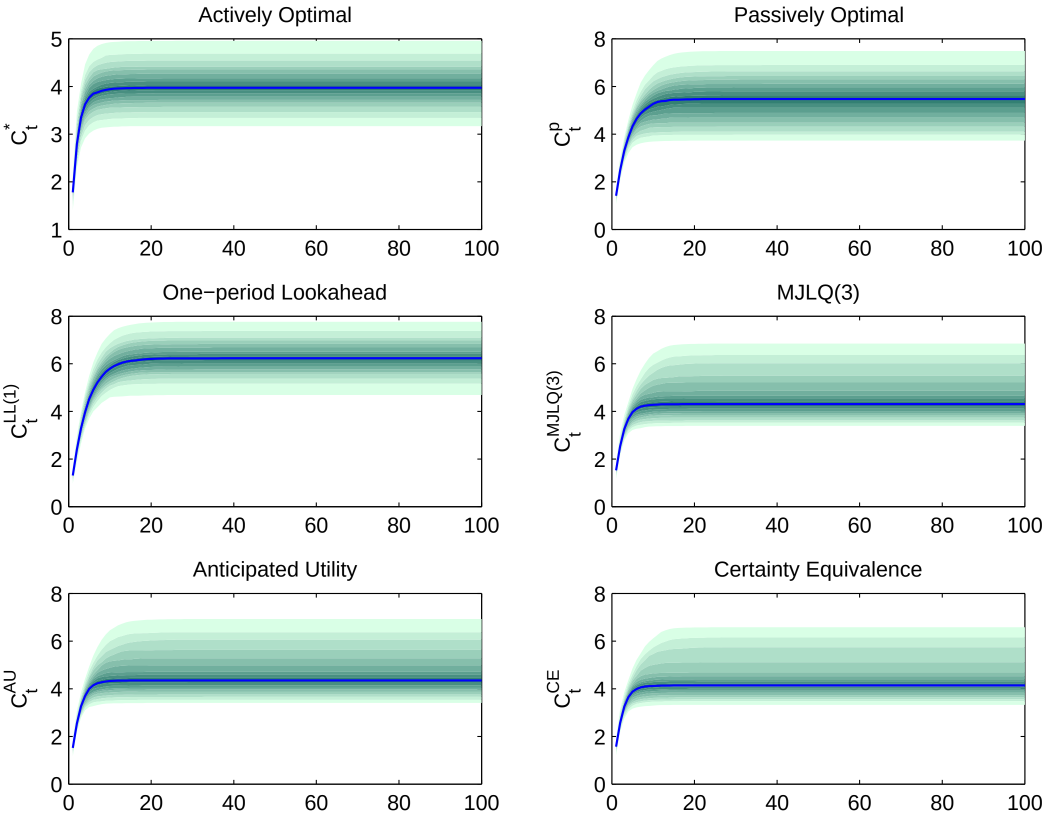

Figure 59: Evolving distribution of simulated time-series of regret function C_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$. Number of simulations: $NMC = 400$.

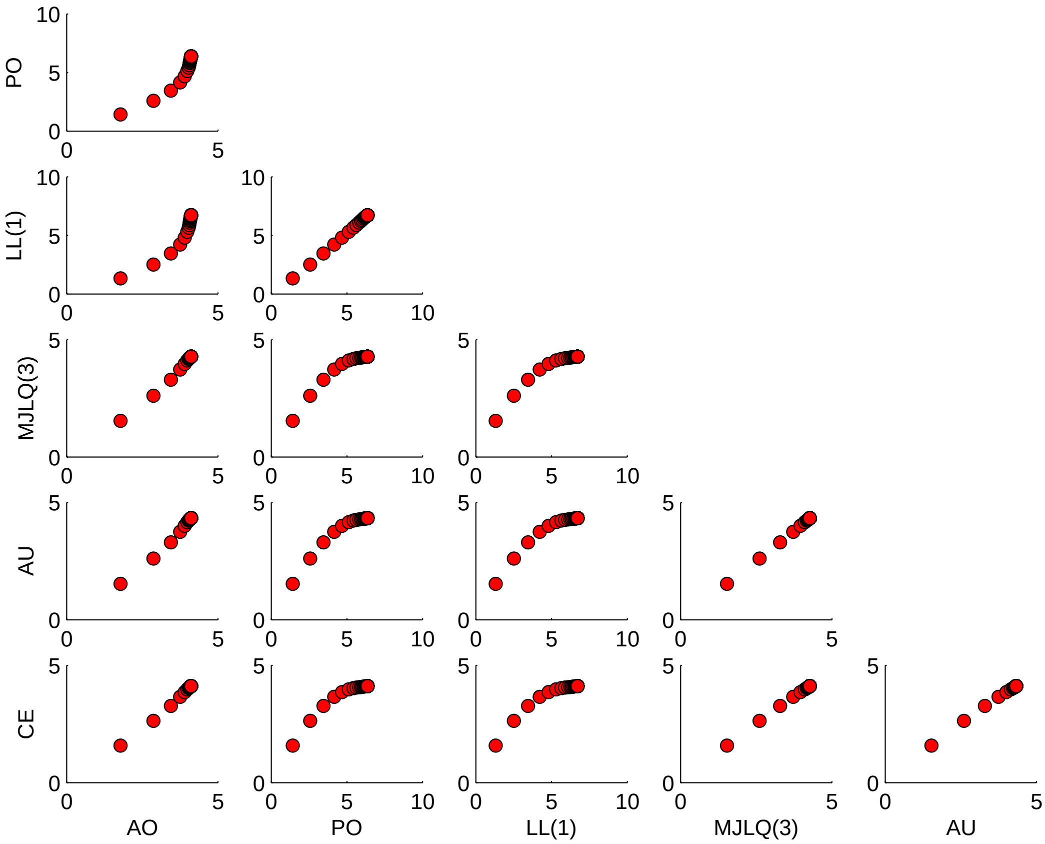

Figure 60: Pairwise comparisons of single simulated time-series of regret function C_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$.

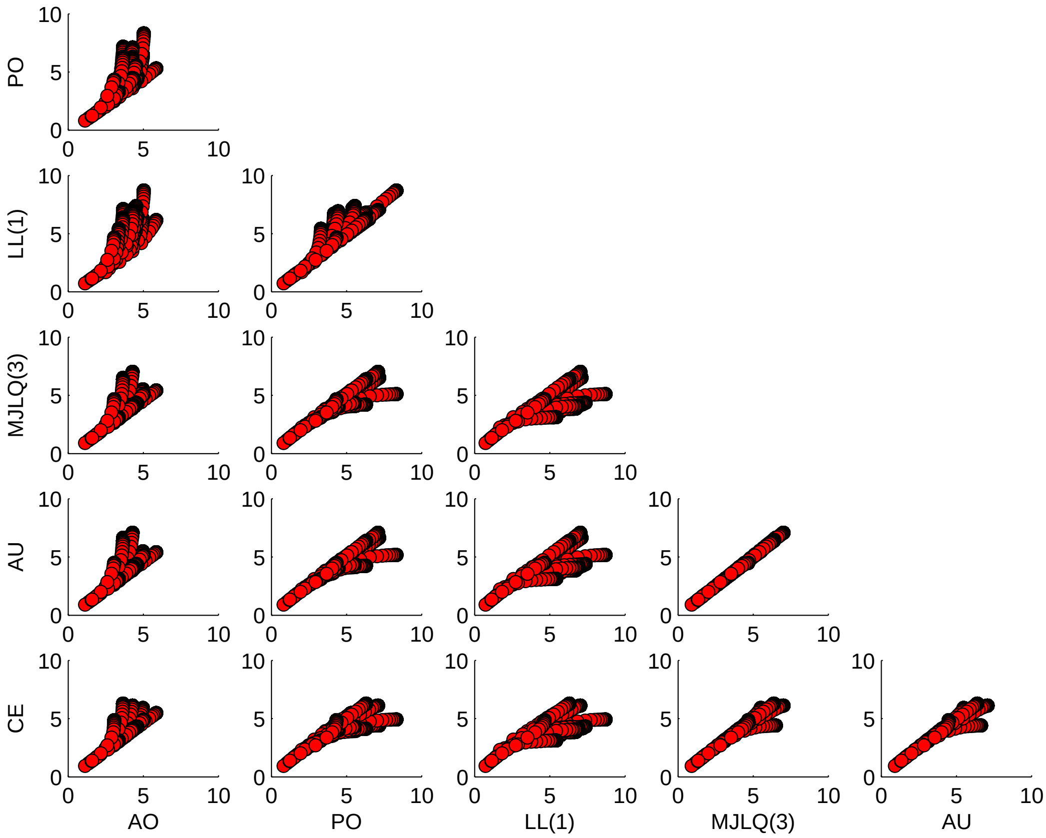

Figure 61: Pairwise comparisons of multiple simulated time-series of regret function C_t under different policies. Parameter values: $\alpha = -0.05$, $\gamma = 0.9$, $\delta = 0.75$, $\sigma_\epsilon^2 = 0.01$, $\sigma_\eta^2 = 0.0001$, $\omega = 1.0$, $x^* = 1.0$, $u^* = 0$. Starting values: $x_0 = 0$, $\mu_0 = -0.1$, $\Sigma_t = 0.04$. True initial slope: $\beta_1 = 0.5$. Number of time periods: $T = 100$. Number of simulations: $NMC = 400$.