LEARNING AND ACTIVE CONTROL OF STATIONARY AUTOREGRESSION WITH UNKNOWN SLOPE AND PERSISTENCE

SERGEI MOROZOV

ABSTRACT. In this study we inquire into the optimal simultaneous Bayesian learning and control of a linear first-order autoregressive stochastic process with unknown policy impact and persistence parameters. Dynamic programming solution to this imperfect information problem has previously been infeasible due to the *curse of dimensionality*. We provide the first characterization of the optimal decision rule that balances exploration Information acquisition) and exploitation (target stabilization). Our numerical results indicate substantial degree of experimentation inherent in the optimal decision rule. Optimal policy is often discontinuous and takes irregular shapes. We identify state space regions where experimentation motive dominates stabilization and visa versa, and explore sensitivity to the model parameters. Second, we contrast the optimal policy against an ensemble of suboptimal alternatives, in hope of identifying key features that could form an arsenal of good rules of thumb to attack higher dimensional problems where dynamic programming is not yet feasible. Our ensemble of approximate solutions includes myopic, certainty equivalent, anticipated utility, limited lookahead policies and assorted hybridizations and modifications, including methods for the actively adaptive prediction of posterior variance. We conclude that aligning the degree of experimentation with that of the optimal policy is essential for good performance of suboptimal approximation.

JEL classification: C44; C63; D83; E17; E52

Keywords: Bayesian dual control, active learning, experimentation, certainty equivalence, anticipated utility, limited lookahead, rules of thumb

It is what we think we know already that often prevents us from learning.
–Claude Bernard

1. INTRODUCTION

This paper studies control of a simple first-order autoregressive process

$$x_t = \beta u_t + \gamma x_{t-1} + \epsilon_t, \tag{1.1}$$

where the slope coefficient β and persistence parameter γ are both unknown with a joint Gaussian prior. Minimizing expected discounted quadratic loss subject to (1.1) is a common stylized representation of many macroeconomic policy problems, such as monetary or fiscal stabilization, exchange rate targeting, pricing of government debt, etc.

The problem belongs to a class of dynamic programs with imperfect information. As the system evolves, new data on the both sides of the regression relationship (1.1) forces revisions of estimates for location and precision that characterize Gaussian posterior beliefs. Anticipation of posterior revisions makes beliefs a part of the state vector, inducing a tradeoff between current stabilization and future sharpening of posterior beliefs. The actively optimal control is the one that balances exploration (information acquisition) and exploitation (target stabilization) optimally. Prescott (1972); Kendrick (1979, 1982); Wieland (2000); **?**); Cogley, Colacito, and Sargent (2007) predict varying degrees of experimentation, depending

Date: September 27, 2007.
Version 0.01.
Preliminary and Incomplete.
Usual disclaimers apply. I thank TEU for providing computer support. All remaining errors are the responsibility of the second author.

on details of the model but they all agree that some degree of experimentation is indeed optimal.

Imperfect information dynamic programs such as this are notoriously hard. Indeed, Bayesian updating is nonlinear, state dimension is high due to the inclusion of beliefs, and cost-to-go function need not be convex with respect to the policy variable. The last problem may imply non-smooth shape of the optimal cost-to-function with respect to the state variables and discontinuities of the optimal policy function. The predicament 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.[1]

The objective of this study is two-fold. First, we provide the first characterization of the actively optimal solution to the simultaneous problem of learning and control that optimally balances exploration (information acquisition) and exploitation (target stabilization) by means of dynamic programming in the state space that is extended to include both physical and informational state variables. Our numerical findings indicate substantial degree of experimentation inherent in the optimal decision rule. Optimal policy is often discontinuous and takes on irregular shape. We identify regions in the state space where the optimizing decision-maker is prompted to explore more actively while foregoing some stabilization performance and regions where stabilization motive is the dominant one. For example, the degree of experimentation decreases as beliefs about persistence approach unit root. In contrast, the experimentation rises with the variance of belief about persistence unless it is so large that the sign of autoregressive coefficient becomes highly uncertain, so that further increases lead now to a reduction of the policy activism. We also explore sensitivity of the solution to the changes in the known model parameters. For instance, we find that the variance of state shock tends to make the optimal policy more cautious, except for some outlying regions in the belief space.

Second, we contrast the features of the optimal control against an ensemble of suboptimal alternatives, in hope of identifying key strategies that help mimic the performance of the actively optimal solution without mounting computational complexity of exhaustive elaboration. These could potentially form an arsenal of good rules of thumb to attack problems of higher dimension where dynamic programming is not yet feasible. Our ensemble of approximate solutions includes myopic control, certainty equivalent rule, anticipated utility policy, limited lookahead and assorted novel hybridizations and modifications, such as methods for the actively adaptive prediction of posterior variance. The comparison is done in terms of the analysis of policy and value functions as well as in terms of simulated Monte Carlo dynamics. We conclude that aligning the degree of experimentation, whether intentional or not, with that of the optimal policy is essential for good performance of suboptimal approximation. In this regard, the policy that incorporates unscented Kalman filter to project future beliefs in the indirect approximate limited lookahead often performs closest to the optimum.

The paper is laid out as follows. Section 2 places our contribution in the context of existing literature. Section 3 sets up the imperfect information dynamic control problem and develops optimal solution by means of dynamic programming. A number of approximate policies is obtained in the section 4. Section 5 is devoted to comparative analysis of various suboptimal policies and their relationships to the fully optimal solution. Section 7 offers concluding remarks. Technical details are relegated to the appendix.

[1]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 is a subject of ongoing research.

2. Review of the Literature

The problem of conflict between information gathering and control quality was originally introduced and discussed by A. A. Feldbaum in a sequence of four seminal papers from 1960 and 1961 (Feldbaum, 1960a,b, 1961a,b). The compromise between probing and control or in Feldbaum's terminology, investigating and directing lead to the concept of dual control. Feldbaum was the first to show that, in principle, the optimal solution can be found by dynamic programming, via what later became known as Bellman functional equation.

A seminal paper by Prescott (1972) introduced active learning into economics, even though it only considered a toy multi-period control problem with the data generated by the simple regression with an unknown slope, $x_t = \beta u_t + \epsilon_t$. Assuming linear quadratic Bayes risk with the entire weight given to deviation of y_t from its target, he solved for the optimal learning policies as functions of beliefs for several small values of the planning horizon (up to 6). The results showed little difference between the myopic and optimal policies except under very large parameter uncertainty. His numerical study also showed the myopic policy to be superior to the certainty equivalent policy in approximating the optimal policy.

The problem of active learning has appeared in many applied economic studies. These include the works of Rausser and Freebairn (1974a,b) on agricultural trade policy, Rothschild (1974); Balvers and Cosimano (1990); Trefler (1993) on monopolistic pricing with unknown demand, Chong and Cheng (1975) on pricing strategies for the introduction of a new product, Bergemann and Välimäki (1996) on strategic pricing, Craine (1979); Bertocchi and Spagat (1993); Yetman (2000); Ellison and Valla (2001); Ellison (2006); Ellison, Sarno, and Vilmunen (2006); Cogley, Colacito, and Sargent (2007) on the optimal monetary policy, Bertocchi (1993) on a theory of floating public debt issues using "subscription issues" when demand schedule for bonds is unknown, Zampolli (2005) on the exchange rate stabilization, Moscarini and Smith (2001) on R&D investment, Hong and Rady (2002) on asset pricing with uncertain supply of liquidity, small scale macroeconometric models of Kendrick (1979); Bar-Shalom and Wall (1980); Kendrick (1982); Amman and Kendrick (1994, 1997) and many others.

3. Actively Optimal Policy

The decision-maker is minimizes discounted intertemporal cost-to-go function with quadratic per-period losses,

$$\min_{\{u_t\}_{t=0}^{\infty}} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \delta^t \left((x_t - x^*)^2 + \omega (u_t - u^*)^2 \right) \right], \tag{3.1}$$

subject to the evolution law of policy target x_t from a class of linear first-order autoregressive stochastic processes[2]

$$x_t = \beta u_t + \gamma x_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2). \tag{3.2}$$

$\delta \in [0, 1)$ is discount factor, x^* is the stabilization target, u^* is "costless" control, $\omega \geq 0$ gives weight to the deviations of u_t from u^*. Variance of the shock, σ_ϵ^2 is known.

The two parameters that govern the conditional mean of x_t, namely the slope coefficient β and persistence γ are unknown. Initially, prior beliefs about $(\beta, \gamma)'$ are jointly Gaussian:

$$\begin{pmatrix} \beta \\ \gamma \end{pmatrix} \sim \mathcal{N}\left(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0\right). \tag{3.3}$$

Gaussian prior (3.3) combined with normal likelihood (3.2) yields Gaussian posterior (Judge, Lee, and Hill, 1988), and so at each point in time the beliefs about unknown parameters β and γ are conditionally normal and are completely characterized by mean vector $\boldsymbol{\mu}_t$ and

[2]Omitting known constant term is without significant loss of generality as target state x^* captures that level effect already. Allowing *unknown* constant term, on the other hand, makes a problem much more complex.

covariance matrix $\mathbf{\Sigma}_t$ (sufficient statistics). Upon observing a realization of x_t, these are updated in accordance with the Bayes law:[3]

$$\mathbf{\Sigma}_{t+1} = \left(\mathbf{\Sigma}_t^{-1} + \frac{1}{\sigma_\epsilon^2}\begin{pmatrix} u_t \\ x_{t-1}\end{pmatrix}\begin{pmatrix} u_t \\ x_{t-1}\end{pmatrix}'\right)^{-1}, \qquad (3.4)$$
$$\boldsymbol{\mu}_{t+1} = \mathbf{\Sigma}_{t+1}\left(\frac{1}{\sigma_\epsilon^2}\begin{pmatrix} u_t \\ x_{t-1}\end{pmatrix} x_t + \mathbf{\Sigma}_t^{-1}\boldsymbol{\mu}_t\right).$$

Note that the evolution of the covariance matrix is not deterministic. Not even a single element of the matrix evolves deterministically. This matters for the approximating dual control methods that try to exploit forecasts of future beliefs in order to achieve outcomes that are closer to the optimal while trying to avoid difficult dynamic programming approach. From (3.4), only one step ahead variance of beliefs could be forecasted deterministically, as x_{t-1} is already known and u_t is fully under control.

Under distributional assumptions (3.2) and (3.3), the imperfect information problem is transformed into the state-space form by defining extended state containing both physical and informational components:[4]

$$S_t = (x_t, \mu_t, \mathrm{vech}(\Sigma_t)). \qquad (3.5)$$

3.1. **Dynamic Programming Formulation.** The stationary Bellman equation is given by

$$V(S_t) = \min_{\{u_{t+1}\}}\Big\{L(S_t, u_{t+1}) + \delta\int V\left(B(S_t, \beta u_{t+1} + \gamma x_t + \epsilon_{t+1}, u_{t+1})\right) p(\beta,\gamma|S_t) q(\epsilon_{t+1}) d\beta d\gamma d\epsilon_{t+1}\Big\}, \qquad (3.6)$$

where $L(S_t, u_{t+1})$ is the expected one-period loss function:
(3.7)

$$\begin{aligned} L(S_t, u_{t+1}) &= \int \left((\beta u_{t+1} + \gamma x_t + \epsilon_{t+1} - x^*)^2 + \omega(u_{t+1} - u^*)^2\right) p(\beta,\gamma|S_t) q(\epsilon_{t+1}) d\beta d\gamma d\epsilon_{t+1} \\ &= \left(\Sigma_{t+1}^{11} + \mu_{1t+1}^2 + \omega\right) u_{t+1}^2 + 2\left((\Sigma_{t+1}^{12} + \mu_{1t+1}\mu_{2t+1})x_t - \mu_{1t+1}x^* - \omega u^*\right) u_{t+1} \\ &+ \omega(u^*)^2 + (\Sigma_{t+1}^{22} + \mu_{2t+1}^2)x_t^2 + (x^*)^2 + \sigma_\epsilon^2 - 2\mu_{2t+1}x^* x_t, \end{aligned}$$

and $p(\beta, \gamma|S_t)$, $q(\epsilon_t)$ represent posterior belief density and density of state shocks, respectively.

Although the stochastic process under control is linear and the loss function is quadratic, the belief updating equations are non-linear, and hence the dynamic optimization problem is more difficult than those in the class of linear quadratic problems. Following Easley and Kiefer (1988), it could be shown that Bellman functional operator is a contraction and a stationary optimal policy exists such that corresponding value function is continuous and satisfies the above Bellman equation. Accordingly, the optimal policy and value functions can be obtained by numerical dynamic programming methods. In particular, we use a combination of the value and policy iterations on six-dimensional grid in the state-space, resulting in the hybrid dynamic programming algorithm. 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

[3]We use subscript $t+1$ to denote beliefs after x_t is realized but before the choice of u_{t+1} is made at the beginning of period $t+1$. This notational timing convention accords with that in Wieland (2000). Technically, it means that u_{t+1} is measurable with respect to filtration $\mathcal{F}_t$ generated by histories of stochastic process up until time t.

[4]vech is a half-vectorization operator, designed to extract non-redundant entries of symmetric matrix. Here, $\mathrm{vech}(\Sigma_t) = \left(\Sigma_t^{11}, \Sigma_t^{12}, \Sigma_t^{22}\right)'$.

evaluation when $\delta = 0.75$. Since the integration step in (3.6) cannot be carried out analytically, we resort to the help of Gauss-Hermite quadrature and multi-linear interpolation. Actively optimal policy and cost-to-go functions are thus 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 six-dimensional hyper-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 hyper-cube remain there for the entire time span of a simulation. The dynamic programming algorithm was iterated to convergence with relative tolerance of $1e - 4$.

3.2. **Features of Actively Optimal Policy.** Results of MacRae (1972) and MacRae (1975), using adaptive covariance actively adaptive methods, suggest that when both the policy impact and state persistence are both unknown the behavior of optimal dual control could be rather complex with respect to parameter uncertainty, correlation of beliefs, and current physical state, in comparison with finite impulse response models. Indeed, when the autoregressive coefficient is not known, the impulse response dynamics can induce potentially long-lasting effects. Not knowing persistence could lead to costly, nearly permanent repercussions. At the same time, cost of experimentation could also be amplified. Also, if the degree of correlation between parameters associated with current control and with lag stages is high, there's significant information to be gained about the two simultaneously and hence experimentation is more advantageous even in the face of large uncertainty about the two parameters by themselves. In this section we summarize the observations we made about the actively optimal policy.

Figure 1 displays one of the (x_t, μ_{1t}) slices of the 6-dimensional optimal control function, while figure 2 displays a representative slice in (x_t, μ_{2t}) plane. Both graphs clearly show that active policy is close to discontinuous and takes irregular shapes with experimentation regions

3.2.1. *Activism and Gradualism.* Nonlinearity of optimal monetary policy that we observed is due to a form of informational endogeneity. As such, it is not limited to Bayesian dual control. For example, it was also found in Tillmann (2008) under min-max strategy to deal with uncertain slope of the Phillips curve. There, informational endogeneity arises through reaction of the worst-case perception of the Phillips curve slope to deviations of inflation away from target. As a result, while small inflation shocks induce cautious policy adjustment, large deviations are combatted vigorously. Empirical studies by Dolado, Pedrero, and Ruge-Murcia (2004); Dolado, Pedrero, and Naveira (2005); Kim, Osborne, and Sensier (2005); Tillmann (2008) find some supporting evidence of nonlinearity in the policy rules.

3.3. **Sensitivity to Model Parameters.** The decision-maker's problem is contains four known parameters – the discount factor δ, the physical shock variance σ_ϵ^2, the target value of the endogenous process x^* and the bliss value of control u^*. In this section we explore the sensitivity of various solutions to these four variables.

3.3.1. *Sensitivity to Discount Factor δ.* Raising the discount factor δ enhances incentive to experiment by making current losses due to experimentation seemingly less important. Accordingly, we observed uniform increase in experimentation by bumping δ up slightly.

3.3.2. *Sensitivity to Physical Shock Variance σ_ϵ^2.* Of the four parameters, the physical shock variance is the most interesting at least for two reasons. First, its impact is intuitively ambiguous because more violent shocks make learning faster but impede progression of endogenous state toward the target. The ambiguity should more pronounced in the regions of divergence between actively optimal and certainty equivalent policies. Second, σ_ϵ^2 is the only parameter in the model not related to preferences. The assumption of perfect

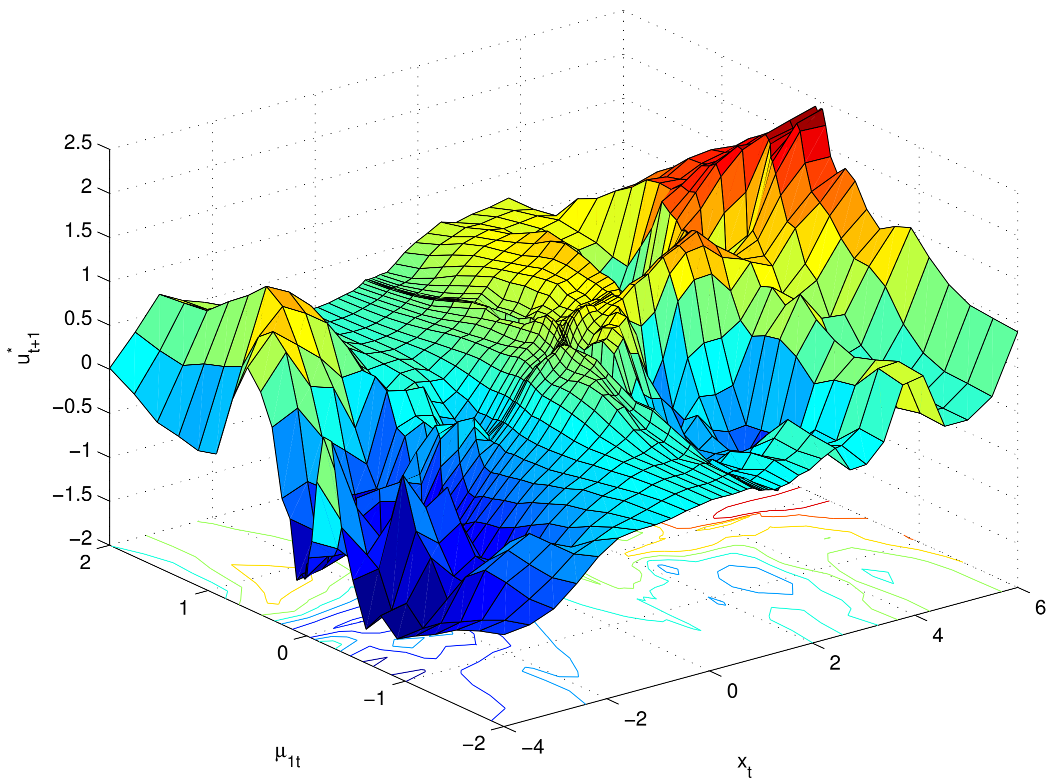

Figure 1: Actively optimal control as a function of x_t and μ_{1t}. Other state coordinates $\mu_{1t} = \rho_t = 0$, $\Sigma_t^{11} = \Sigma_t^{22} = 1.0$. Parameters: $\delta = 0.75$, $\omega = 1.6$, $u^* = 0$, $x^* = 1$, $\sigma_\epsilon^2 = 0.04$.

knowledge of state process variance could be questionable and thus the range of values need to be explored. Making σ_ϵ^2 formally unknown to the policy maker is beyond the scope of this paper.[5]

As with δ, we limit our exploration of sensitivity to σ_ϵ^2 to trying out small number of alternative values. Here we found that the variance of the state shock tends to make the optimal policy more cautious, except for some outlying regions in the state space.

4. Ensemble of Approximations

In this section we introduce a large number of approximate solutions. Some of these are popular in extant dual control literature, some are more seldom seen, and some are altogether new.

4.1. **Myopic Certainty Equivalent Control.**

The simplest form of control is to disregard altogether both model uncertainty and dynamic link between the present and the future focusing instead on the minimization of immediate period loss function. Since immediate period loss function involves expectation of target state x_{t+1} in response to the choice of control u_{t+1}, eschewing model uncertainty means replacing unknown coefficients with best current estimates, i.e. the mean posterior beliefs. Resulting myopic certainty equivalent policy is given by

$$u_{t+1}^{MCE} = -\frac{\mu_{1t+1}\mu_{2t+1}}{\mu_{1t+1}^2 + \omega} x_t + \frac{\mu_{1t+1}x^* + \omega u^*}{\mu_{1t+1}^2 + \omega}. \tag{4.1}$$

[5] Sensitivity exploration could serve to calibrate the prior for the shock variance in a way to best match 'stylized facts' (Canova, 2007), in an empirical Bayes spirit.

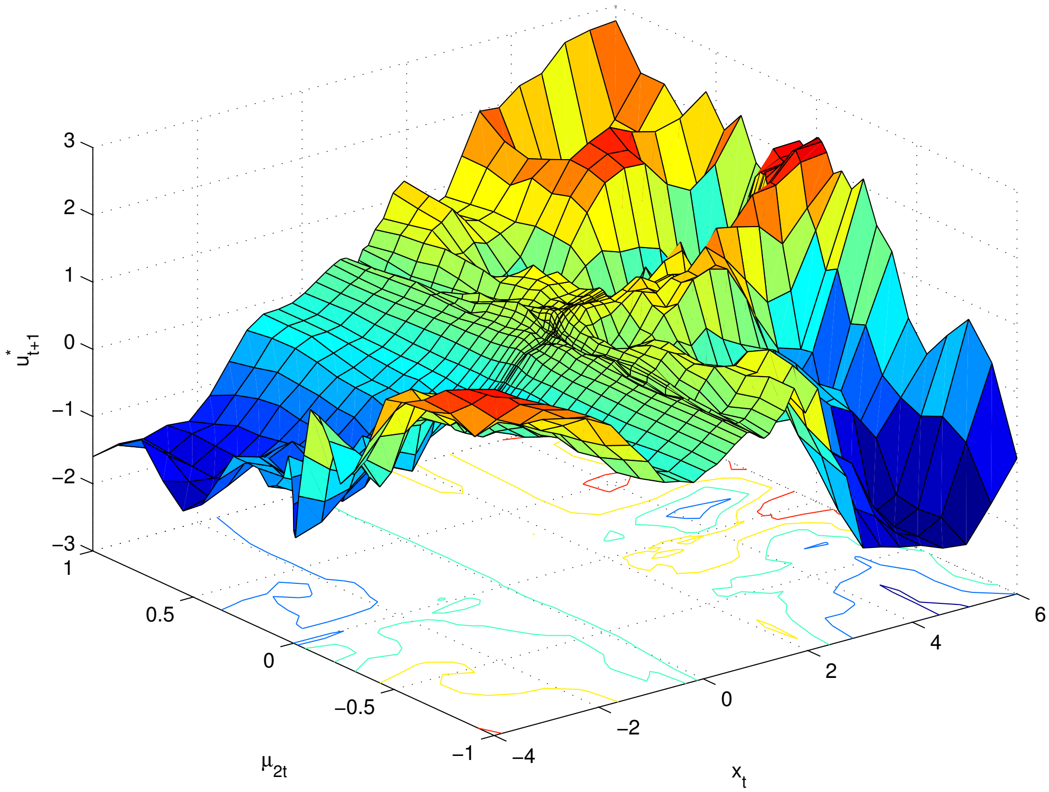

Figure 2: Actively optimal control as a function of x_t and μ_t^1. Other state coordinates $\mu_t^2 = \rho_t = 0$, $\Sigma_t^{11} = \Sigma_t^{22} = 1.0$. Parameters: $\delta = 0.75$, $\omega = 1.6$, $u^* = 0$, $x^* = 1$, $\sigma_\epsilon^2 = 0.04$.

Under certainty equivalent myopic policy

$$\mathbb{E}_t^{MCE}(x_{t+1}) = \frac{\mu_{2t+1}\omega}{\mu_{1t+1}^2 + \omega} x_t + \frac{\mu_{1t+1}(\mu_{1t+1}x^* + \omega u^*)}{\mu_{1t+1}^2 + \omega}. \tag{4.2}$$

If control itself is costless, i.e. $\omega = 0$, myopic certainty equivalent policy displays no gradualism and the state is expected to hit the target x^* in one period.

4.2. **Cautionary Myopic Control.** The second type of myopic policy recognizes that the state transition parameters are random variables distributed normally with mean and variance derived from Bayesian learning. We call such policy *cautionary myopic*, as is traditional in the dual control literature. The rule is myopic in that it ignores the link between current and future state and beliefs. Policy myopia has been extensively studied. In the context of adaptive learning and control in economics, Prescott (1972) is an early example that found little difference between the optimal and myopic policy. Cautionary myopic policy also results as a first iterate of the value function iteration algorithm. Therefore, knowing the myopic policy explicitly allows to skip one step.

Cautionary myopic control solves myopic control problem

$$\min_{u_{t+1}} \left\{ \mathbb{E}_t \left(\beta u_{t+1} + \gamma x_t - x^* + \epsilon_{t+1}\right)^2 + \omega(u_{t+1} - u^*)^2 \right\}, \tag{4.3}$$

where vector $(\beta, \gamma)'$ is treated according to predictive distribution, i.e. as normally distributed random variate with mean $\boldsymbol{\mu}_{t+1|t}$ and covariance matrix $\boldsymbol{\Sigma}_{t+1|t}$. Straightforward calculation shows that the solution is

$$u_{t+1}^{MYOP} = -\frac{(\Sigma_{t+1}^{12} + \mu_{1t+1}\mu_{2t+1})}{\Sigma_{t+1}^{11} + \mu_{1t+1}^2 + \omega} x_t + \frac{\mu_{1t+1}x^* + \omega u^*}{\Sigma_{t+1}^{11} + \mu_{1t+1}^2 + \omega}. \tag{4.4}$$

Implied expected dynamics of the state under control (4.4) becomes

$$\mathbb{E}_t^{MYOP}(x_{t+1}) = \frac{\mu_{2t+1}\Sigma_{t+1}^{11} - \mu_{1t+1}\Sigma_{t+1}^{12} + \omega\mu_{2t+1}}{\Sigma_{t+1}^{11} + \mu_{1t+1}^2 + \omega} x_t + \frac{\mu_{1t+1}(\mu_{1t+1}x^* + \omega u^*)}{\Sigma_{t+1}^{11} + \mu_{1t+1}^2 + \omega}. \tag{4.5}$$

4.3. **Certainty Equivalent Solution.** The certainty equivalent solution corresponds to a strategy of disregarding uncertainty embodied in the posterior distribution of regression coefficients but instead solving intertemporal optimization problem with parameters set at their mean values. After new state realization is observed the parameter estimates are updated in accordance with the Bayes rule, and new certainty-equivalent policy will be computed. The certainty-equivalent decision maker is therefore moderately schizophrenic in the way uncertain parameters are treated. Parameters of the target state process are treated as random variables when agents learn but as constants when they formulate decisions. Looking backward, agents can see how their beliefs have evolved in the past, but looking forward they act as if they have attained the ultimate beliefs already. Moreover, in a certainty-equivalent world these beliefs collapse all of the probability density of the parameters into a single point – the current mean estimate of the parameters. It also doesn't matter if the additive noise is present or not – the certainty equivalent control is known to be optimal in the case of constant coefficients with or without ϵ_t and is independent of the shock variance. While it constitutes a useful benchmark, it is definitely suboptimal in the case of multiplicative parameter uncertainty.

In a linear-quadratic Gaussian setting, the certainty equivalent control can be calculated by guess-verifying quadratic form of the cost-to-go function. The control is then linear in the physical state x_t:

$$u_t^{CE} = -\frac{\mu_{1t+1}\mu_{2t+1}(1+\delta A)}{\mu_{1t+1}^2 + \omega + \delta A\mu_{1t+1}^2} x_{t-1} + \frac{\mu_{1t+1}x^* + \omega u^* - \delta\mu_{1t+1}B}{\mu_{1t+1}^2 + \omega + \delta A\mu_{1t+1}^2}, \tag{4.6}$$

where A and B are the first two coefficients in the quadratic representation of the optimal cost function:

$$V^{CE}(x_t) = Ax_t^2 + 2Bx_t + C. \tag{4.7}$$

Coefficient A solves one-dimensional version of algebraic Riccati equation (Hansen and Sargent, 2004), which here is simply a larger positive root of the quadratic equation

$$\mu_{2t+1}^2\omega(1+\delta A) - A\left(\mu_{1t+1}^2(1+\delta A) + \omega\right) = 0, \tag{4.8}$$

while B is related to A via the following equation:

$$B = \frac{\omega\left(\mu_{1t+1}\mu_{2t+1}u^* + \mu_{1t+1}\mu_{2t+1}\delta A u^* - \mu_{2t+1}x^*\right)}{\mu_{1t+1}^2 + \omega + \delta A\mu_{1t+1}^2 - \mu_{2t+1}\delta\omega}. \tag{4.9}$$

If $\omega = 0$, the certainty equivalent control results in expected stabilization in one period. $\delta = 0$ results in certainty equivalent myopic control (4.1). Given mean beliefs $\boldsymbol{\mu}$, coefficients A, B and C are treated as constants (hence omission of time subscripts), yet these are recalculated anew at all belief updates.

4.4. **Anticipated Utility Policy.** Anticipated utility decision-maker behaves similar to the certainty equivalent agent in the way the choices are made and beliefs are updated. The only distinction is recognition of uncertainty in formulation of decisions. Unknown parameters are treated as random variables with distribution that is fixed over time.

It can be shown that anticipated utility policy remains linear in the observed state:

$$u_{t+1}^{AU} = -\frac{(\Sigma_{t+1}^{12} + \mu_{1t+1}\mu_{2t+1})(1+\delta A)}{(\Sigma_{t+1}^{11} + \mu_{1t+1}^2)(1+\delta A) + \omega} x_t + \frac{\mu_{1t+1}x^* + \omega u^* - \mu_{1t+1}\delta B}{(\Sigma_{t+1}^{11} + \mu_{1t+1}^2)(1+\delta A) + \omega}, \tag{4.10}$$

where A and B are the highest order coefficients in the quadratic representation of the optimal cost function

$$V^{AU}(x_t) = Ax_t^2 + 2Bx_t + C. \tag{4.11}$$

Coefficient A solves one-dimensional Bayesian linear regulator (Hansen and Sargent, 2004), i.e. the largest positive root of the quadratic equation[6]

$$(1+\delta A)^2\left((\Sigma^{11}_{t+1}+\mu^2_{1t+1})(\Sigma^{22}_{t+1}+\mu^2_{2t+1})-(\Sigma^{12}_{t+1}+\mu_{1t+1}\mu_{2t+1})^2\right) \\ +(1+\delta A)\left(\omega(\Sigma^{22}_{t+1}+\mu^2_{2t+1})-A(\Sigma^{11}_{t+1}+\mu^2_{1t+1})\right)-A\omega=0, \tag{4.12}$$

while B is related to A via the following equation:

$$\frac{(\Sigma^{12}_{t+1}+\mu_{1t+1}\mu_{2t+1})(1+\delta A)\left(\mu_{1t+1}x^*+\omega u^*-\mu_{1t+1}\delta B\right)}{(\Sigma^{11}_{t+1}+\mu^2_{1t+1})(1+\delta A)+\omega}+\delta\mu_{2t+1}B-\mu_{2t+1}x^*=B. \tag{4.13}$$

Even with $\omega=0$ anticipated utility control no longer results in perfect one-period stabilization in expectation. Setting $\delta=0$ reduces anticipated utility solution to its myopic special case. Similarly, annihilating belief covariance $\mathbf{\Sigma}_{t+1}=0$ reproduces the certainty equivalent case.

Dynamically, the process unfolds as follows. In a period t, given current beliefs, the anticipated utility agent applies control u^{AU}_{t+1} per equation (4.10). Following the choice of control, next period state is realized according to (3.2), and the beliefs about the uncertain parameters are updated via Bayes law (3.4). The process repeats itself in period $t+1$. The outcomes will be shown in more details in section 5 where we collate results of several alternative policies to ease their comparison.

4.5. **Direct Limited Lookahead with Prediction of Posterior Variance.** Myopic certainty equivalent policy and cautionary myopic rule are completely myopic in that only immediate payoff is considered. In contrast, the dual optimal policy takes full account of future payoffs and intertemporal links. Limited lookahead approaches intermediate between these two extremes by considering only a small number of future periods (Bertsekas, 2005). Limited lookahead can either be spelled out directly in terms of explicit finite sequence of controls that optimize finite horizon criterion, or indirectly via cost-to-go function (i.e. dynamic programming). We consider some direct limited lookahead approaches first.

The n-period limited lookahead control solves

$$\min_{u_{t+1},\ldots,u_{t+n+1}}\Bigg\{\mathbb{E}_t\left(\beta u_{t+1}+\gamma x_t+\epsilon_{t+1}-x^*\right)^2 \\ +\sum_{\tau=1}^{n}\delta^\tau\mathbb{E}_t\left[(x_{t+\tau+1}-x^*)^2\,\middle|u_{t+1},\ldots,u_{t+\tau}\right]+\omega\sum_{\tau=0}^{n}\delta^\tau\left(u_{t+\tau+1}-u^*\right)^2\Bigg\}, \tag{4.14}$$

subject to the state equation (3.2).

The constituents of the second sum could be expanded by recursive substitution which shows the necessity to predict future moments: $\mathbb{E}_t\left[\gamma^2\middle|u_{t+1},\ldots,u_{t+\tau}\right]$, $\mathbb{E}_t\left[\gamma^4\middle|u_{t+1},\ldots,u_{t+\tau}\right]$, $\ldots$, $\mathbb{E}_t\left[\gamma^{2(\tau+1)}\middle|u_{t+1},\ldots,u_{t+\tau}\right]$, $\mathbb{E}_t\left[\gamma^{\tau+1}\middle|u_{t+1},\ldots,u_{t+\tau}\right]$, $\mathbb{E}_t\left[\beta\gamma\middle|u_{t+1},\ldots,u_{t+\tau}\right]$, $\ldots$, $\mathbb{E}_t\left[\beta\gamma^{2\tau+1}\middle|u_{t+1},\ldots,u_{t+\tau}\right]$, $\mathbb{E}_t\left[\beta\middle|u_{t+1},\ldots,u_{t+\tau}\right]$, $\mathbb{E}_t\left[\beta^2\middle|u_{t+1},\ldots,u_{t+\tau}\right]$, $\mathbb{E}_t\left[\beta^2\gamma\middle|u_{t+1},\ldots,u_{t+\tau}\right]$, $\ldots$, $\mathbb{E}_t\left[\beta^2\gamma^{2\tau}\middle|u_{t+1},\ldots,u_{t+\tau}\right]$, for all $\tau=1,\ldots,n$. Considerable difficulty here is caused by the fact that while the vector $(\beta,\gamma)'$ is jointly normal conditional on the past history of both target states and control up to date τ, i.e. given $\{(x_{t+s},u_{t+s})\}^\tau_{s=0}$, marginalizing $\{x_{t+s}\}^\tau_{s=0}$ is hard. This is because integrating future target state observations $\{x_{t+s}\}^\tau_{s=0}$ out of τ-period ahead Bayesian updating equations is a nonlinear operation.

Instead of integrating future observations out, the first approximate solution that I dub "limited lookahead prediction by mean values" (LLPMV) replaces future random x_{t+s} values by their expected values conditional on future control sequence $\{u_{t+s}\}^\tau_{s=1}$. In other words, the nonlinear Bayesian updating of future information states is constrained to the conditional

[6]Positivity of the larger root is not guaranteed, in contrast to the certainty-equivalent case.

nominal path:

(4.15)

$$\widehat{x}_{t+s}=\mathbb{E}_t\left(\beta\middle|u_{t+1},\ldots,u_{t+\tau}\right)\sum_{j=1}^{s}u_{t+j}+\sum_{j=1}^{s}\left(\mathbb{E}_t\left(\gamma\middle|u_{t+1},\ldots,u_{t+\tau}\right)\right)^j\widehat{x}_{t+s-j},\quad s=1,\ldots,\tau-1.$$

For simplicity, we consider two special cases $n = 1$ and $n = 2$. In the former case, we need to compute $\mathbb{E}_t\left(\beta\middle|u_{t+1}\right)$, $\mathbb{E}_t\left(\gamma\middle|u_{t+1}\right)$, $\mathbb{E}_t\left(\beta^2\middle|u_{t+1}\right)$, $\mathbb{E}_t\left(\gamma^2\middle|u_{t+1}\right)$, $\mathbb{E}_t\left(\beta\gamma\middle|u_{t+1}\right)$, $\mathbb{E}_t\left(\beta\gamma^2\middle|u_{t+1}\right)$, $\mathbb{E}_t\left(\beta^2\gamma\middle|u_{t+1}\right)$, $\mathbb{E}_t\left(\beta^2\gamma^2\middle|u_{t+1}\right)$, $\mathbb{E}_t\left(\beta\gamma^3\middle|u_{t+1}\right)$, $\mathbb{E}_t\left(\gamma^4\middle|u_{t+1}\right)$. In turn, these are the moments (including higher order ones) of $p(\beta,\gamma|u_{t+1})=\int_{x_t}p(\beta,\gamma|u_{t+1},x_t)dx_t$. Since x_t is known, integration is redundant. In the latter case, we also require $\mathbb{E}_t\left(\gamma^3\middle|u_{t+1},u_{t+2}\right)$, $\mathbb{E}_t\left(\beta^2\gamma^3\middle|u_{t+1},u_{t+2}\right)$, $\mathbb{E}_t\left(\beta\gamma^4\middle|u_{t+1},u_{t+2}\right)$, $\mathbb{E}_t\left(\beta^2\gamma^4\middle|u_{t+1},u_{t+2}\right)$, after replacing random future covariance $\mathbf{\Sigma}_{t+3}$ with its nominal path value, and extending the first set of moments to an analogous set conditional on both u_{t+1} and u_{t+2}. These are all high order moments of multivariate normal distribution whose analytic but tedious representations could be found in (Pearson and Young, 1918; Kendall and Stuart, 1963; Holmquist, 1988, 1996; Kotz, Balakrishnan, and Johnson, 2000; Triantafyllopoulos, 2002). It turns out that under LLPMV(1) and LLPMV(2) formulations, the objective function could be non-convex and have multiple local minima. Multiple optima of this kind are symptomatic of the non-convexity of the dual control problem (Kendrick, 1978; Radner and Stiglitz, 1984; Mizrach, 1991) and call for safeguarded optimization methods such as Nelder-Meade polytope method with multiple random starts, direct search, genetic algorithms or simulated annealing. For robustness we used several of these to ensure the correct optimum is selected at some significant computational expense. Multi-period limited lookahead policies with prediction by mean values could also be developed. Unfortunately, the required algebra gets rather involved, even with the help of computer algebra systems (`Mathematica`, `Maple`, etc.). Recursive algorithms to generate high-order moments of bivariate normal distribution do exist (Triantafyllopoulos, 2002), yet coding quickly becomes tedious and error-prone. Complexity and fragility of dense multi-dimensional optimization also grows rapidly with the number of periods under control.[7] Furthermore, it yields no more insight than comparing one-period limited lookahead, two-period limited lookahead with prediction by mean values and fully optimal policy. Instead, it is of more interest to study simplified rather than more complicated solutions. This complexity consideration limits the size of the direct lookahead to no more than two.

The second approximate solution tries to be more accurate by using the second order approximation to the future belief dynamics, i.e. using second order expansion for the mean of the nonlinear function of a random variable. I call this method "limited lookahead prediction by variance" (LLPV). For the one-period lookahead this is not needed as the dynamics of posterior variance could be predicted exactly. For the two-period lookahead we need to compute the expectation of $\mathbf{\Sigma}_{t+3}$ from a second order expansion of the two-fold application of Bayes updating. The contribution of second order terms could be substantial in relative terms. So substantial, in fact, that the resulting approximation to the covariance matrix need not even be positive definite because our direct parametrization of posterior covariance matrix in makes no use of positive definiteness property. While there are parameterizations that enforce this property globally, they involve additional nonlinear transformations. Figure 3 explores the question of gradualism of the LLPV(2) policy. This policy is evidently fairly aggressive in that $\mathbb{E}_t^{LLPV(2)}(x_{t+1})$ is closer to x^* than many less forward-looking approximate policies for which expected target dynamics is closer to the diagonal as we'll see later . The expected profile takes particularly strange and aggressive shape when $\omega = 0$. Tracing out cobweb dynamics against that curve results in very rapid convergence in expectation to the rest point (at the intersection with the diagonal) especially if x_t is not near local minimum around $x_t = -1$. Overall, highly nonlinear shape of the expected state evolution

[7]It could be shown that solving general n-period lookahead is equivalent to minimizing multivariate polynomial of degree $2(n+1)^2+2$.

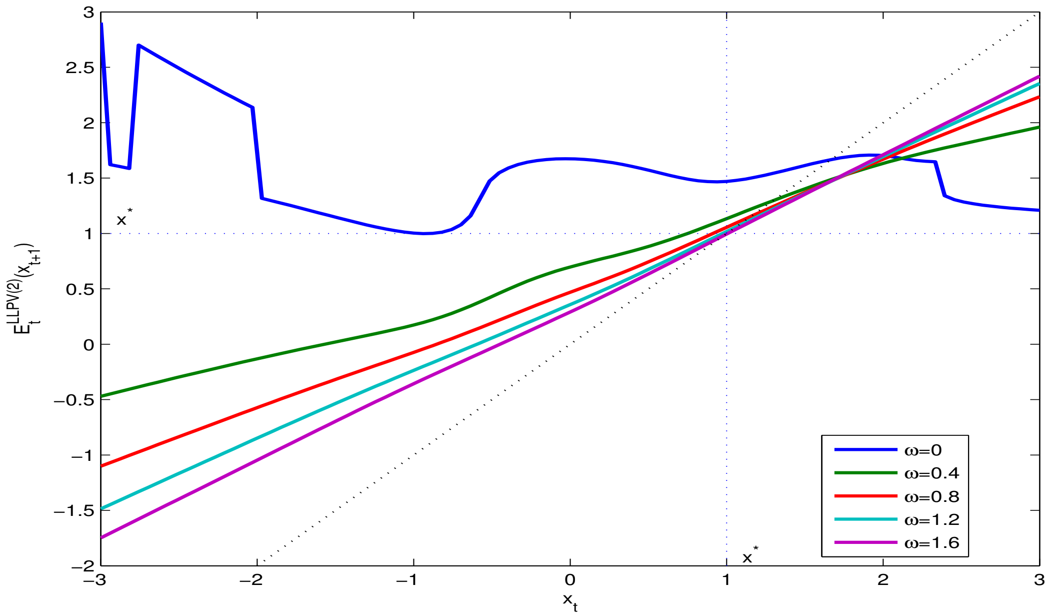

Figure 3: Expected target evolution under modified two period limited lookahead control with second order prediction of the two-step ahead posterior covariance matrix (LLPV(2)). Beliefs: $\boldsymbol{\mu}_{t+1} = (-0.5, 0.9)'$, $\Sigma^{11}_{t+1} = 0.04$, $\Sigma^{12}_{t+1} = 0.0$, $\Sigma^{22}_{t+1} = 0.04$. Parameter values: $\delta = 0.75$, $x^* = 1$, $u^* = 0$, $\sigma^2_\epsilon = 0.04$.

function reflects complicated tradeoffs involved in second order approximation of two-step ahead posterior covariance matrix, learning and control. Indeed, note that minimizing the loss criterion under the second order approximation to the two-step ahead posterior covariance matrix could be reduced to minimization of trivariate polynomial of degree 71. This suggests that developing higher order approximations or expanding the lookahead horizon is not only will be extremely cumbersome but will also fall victim to the usual vagaries of highly nonlinear numerical mathematics. For this reason, we do not attempt to explore more accurate approximations here.

4.6. **Indirect Limited Lookahead via Finite Horizon Dynamic Programming.** Next, we turn to limited lookahead formulations based on the dynamic programming. These approaches require some form of the cost-to-go function to be evaluated at any permissible state space vector. As such these approaches are indirect. Since we are looking for approximate solutions here, we can use limited lookahead as a source of approximation ideas.

The most exact formulation is to consider value iteration used to find dual optimal policy and restrict the number of iterations to a preset small value. With tensor product discretization of the state space, this approach is still subject to the curse of dimensionality in terms of storage requirements although computing time may be substantially reduced by performing far fewer cost-to-go function iterations than is needed to achieve convergence.

Figure 4 sheds some light on the impact of the lookahead horizon on the gradual appearance of nonlinear and experimental features of actively optimal policy.

4.7. **EKF-based Indirect Approximate Limited Lookahead.** The finite horizon dynamic programming algorithm outlined in the previous subsection is still computationally onerous due to massive storage requirements. Further simplifications and approximations could be found by reducing the complexity of the cost-to-go function. We could then compensate information loss by possibly increasing lookahead horizon.

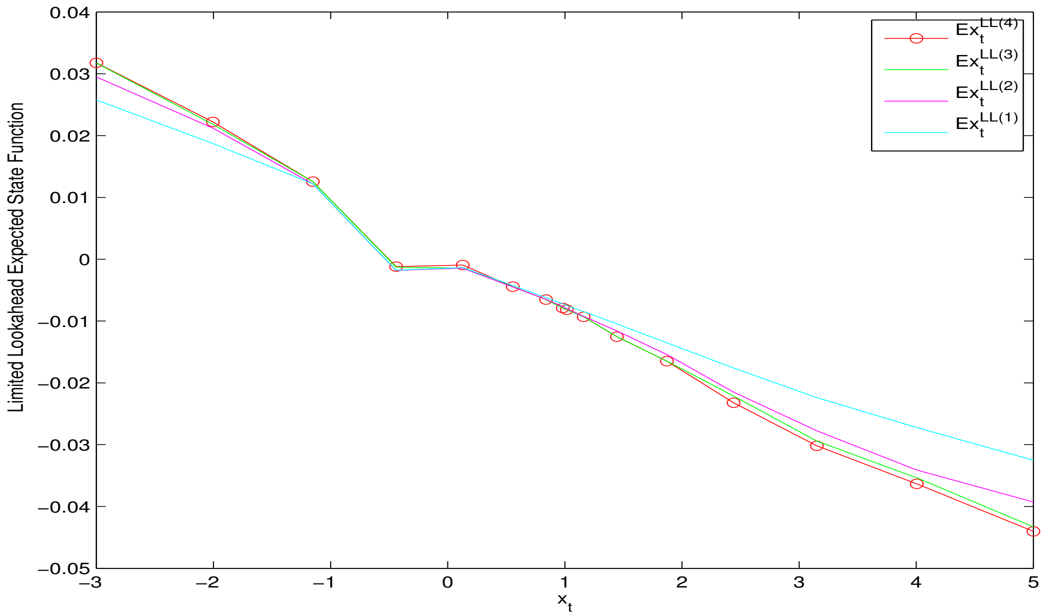

Figure 4: Expected target evolution under various exact limited lookahead controls. Beliefs: $\boldsymbol{\mu}_{t+1} = (-0.5, 0.9)'$, $\Sigma^{11}_{t+1} = 0.04$, $\Sigma^{12}_{t+1} = 0.0$, $\Sigma^{22}_{t+1} = 0.04$. Parameter values: $\delta = 0.75$, $\omega = 0$, $x^* = 1$, $u^* = 0$, $\sigma^2_\epsilon = 0.04$.

One specific idea how to simplify the cost-to-go function was developed in Kendrick (1978) on the basis of Tse, Bar-Shalom, and Meier (1973); Bar-Shalom, Tse, and Larson (1974); Bar-Shalom and Tse (1976) and others. In this approach, we first form a different state space representation by tracking the target state x_t and unobserved coefficients β and γ (augmented state) instead of x_t plus beliefs. Then we formulate a perturbation problem around the certainty equivalent nominal path by substituting second order expansion of Bellman equation into the augmented state dynamics and dropping high order terms. The perturbation problems is approximately from linear quadratic class and could be solved *conditional on future augmented state covariances.* Specifically, perturbation form of augmented state $S_t = (x_t, \beta_{t+1}, \gamma_{t+1})'$ transition around nominal path is

(4.16)

$$dS_{t+1} = \begin{pmatrix} dx_{t+1} \\ d\beta_{t+2} \\ d\gamma_{t+2} \end{pmatrix} = \begin{pmatrix} \gamma_{t+1} & u_{t+1} & x_t \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} dx_t \\ d\beta_{t+1} \\ d\gamma_{t+1} \end{pmatrix} + \begin{pmatrix} \beta_{t+1} \\ 0 \\ 0 \end{pmatrix} du_{t+1} + \begin{pmatrix} d\beta_{t+1} du_{t+1} \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} d\gamma_{t+1} dx_t \\ 0 \\ 0 \end{pmatrix} + \xi_{t+1},$$

and second order expansion of the optimal cost-to-go function

$$J_t(S_t) = J^{CE}_t(S_t) + dJ_t(S_t), \tag{4.17}$$

where J^{CE}_t is the cost-to-go function associated with a nominal path the perturbation cost-to-go function is approximately quadratic:

$$dJ_t(S_t) = \tilde{c}_t + \mathbb{E}_t \left(\tilde{\mathbf{B}}'_t dS_t + \frac{1}{2} dS'_t \tilde{\mathbf{A}}_t dS_t \middle| \mathcal{P}^t \right). \tag{4.18}$$

Recursive expressions for the coefficient matrix $\tilde{\mathbf{A}}_t$, vector $\tilde{\mathbf{B}}_t$ and scalar $\tilde{c}_t$ are computed following Kendrick (2002) assuming finite horizon setup with zero terminal conditions, again, *conditional on future augmented state covariances.* To make the method operational, Kendrick (2002) uses the second order extended Kalman filter whereby state space equations are linearized first and then standard Kalman filter is applied.

Similar to direct lookahead approaches, the approximate optimal cost-to-go function can be split into sum of convex and concave components that complicate its overall shape, including multiple minima, and induce policy function with discontinuities.

4.8. **UKF-based Indirect Approximate Limited Lookahead.** This alternative is an improvement of the preceding section that replaces second order extended Kalman filter with unscented Kalman filter (UKF) of Julier and Uhlmann (1997). The idea of the unscented Kalman filter is that approximation of the posterior density of the state by a Gaussian density is better than linearizing the state transition. It achieves this by propagating deterministic set of specially calibrated points through the true nonlinearity and reconstructing posterior Gaussian density based on the propagated outcomes. In effect, the UKF approximates the first two moments needed for the Kalman update and could be reinterpreted in the direct lookahead framework as a particular integration scheme to integrate out future observations that impact future beliefs. In general, it is accurate to the second order, and to the third order if the prior density was Gaussian.[8]

5. Comparative Analyses

5.1. **Comparing Objects Implied by Policy Functions.**

5.1.1. *Controls.*

5.1.2. *Expected States.*

5.1.3. *Expected Beliefs.*

5.1.4. *Speed of Learning.*

5.2. **Comparing objects Implied by Value Functions.**

5.3. **Comparing Simulated Outcomes.** Priors for simulations are set subjectively.[9]

6. Comparative Sensitivity Analyses

Here we complement the sensitivity analysis of the actively optimal policy with respect to the model parameters by studying impact of model parameters on the approximation errors of various suboptimal alternatives.

7. Concluding Remarks

7.1. **Novel Results.** Compared to the results of Wieland (2000) the actively optimal policy function becomes even more complex.

This results accords with the intuition of Kendrick (1979, 1982) predicting larger role for the experimentation with more sources of uncertainty.

[8]Higher order accuracies for moderately sized lookaheads could be obtained with the Smolyak Kalman filter (Winschel and Krätzig, 2008).

[9]Alternatively, one can use the least informative density consistent with available information from a cross-section of earlier empirical studies in the spirit of El-Gamal (1993) and Canova (1995). I use purely subjective approach to avoid issues of dependence, reliability and compatibility of various parameter estimates available in the literature.

7.2. **Implications for Empirical Work.** Regarding the focus on simultaneous learning and control, our model carries both positive and normative implications for applied work. For example, if the model could be phrased as a positive description of optimal monetary policy process.

If the model is regarded to be a good approximation to the actual data generating process, it can be fruitfully applied to evaluate policy options. Importantly, can actively optimal monetary policy prescribe more or less intentional experimentation above and beyond experimentation embedded in passive learning options such as anticipated utility?

7.3. **Related Research Directions.** Our methods and findings prompt for additional research in several related directions.

References

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

——— (1997): "Erratum to active learning: Monte Carlo results," *Journal of Economic Dynamics and Control*, 21(10), 1613–1614.

Balvers, R. J., and T. F. Cosimano (1990): "Actively Learning About Demand and the Dynamics of Price Adjustment," *Economic Journal*, 100(402), 882–898.

Bar-Shalom, Y., and E. Tse (1976): "Caution, Probing and the Value of Information in the Control of Uncertain Systems," *Annals of Economic and Social Measurement*, 5, 323–336.

Bar-Shalom, Y., E. Tse, and R. E. Larson (1974): "Some Recent Advances in the Development of Closed-Loop Stochastic Control and Resource Allocation Algorithms," in *Proceedings of IFAC Symposium on Stochastic Control*, Budapest, Hungary.

Bar-Shalom, Y., and K. D. Wall (1980): "Dual adaptive control and uncertainty effects in macroeconomic systems optimization," *Automatica*, 16(2), 147–156.

Bergemann, D., and J. Välimäki (1996): "Learning and Strategic Pricing," *Econometrica*, 64(5), 1125–1149.

Bertocchi, G. (1993): "Theory of Public Debt Management with Unobservable Demand," *Economic Journal*, 103(419), 960–974.

Bertocchi, G., and M. Spagat (1993): "Learning, experimentation, and monetary policy," *Journal of Monetary Economics*, 32(1), 169–178.

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

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

Canova, F. (1995): "Sensitivity Analysis and Model Evaluation in Simulated Dynamic General Equilibrium Economies," *International Economic Review*, 36(2), 477–501.

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

Chong, C. Y., and D. C. Cheng (1975): "Multistage pricing under Uncertain Demand," *Annals of Economic and Social Measurement*, 4, 311–324.

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.

Craine, R. (1979): "Optimal Monetary Policy with Uncertainty," *Journal of Economic Dynamics and Control*, 1(1), 59–83.

Dolado, J. J., R. M.-D. Pedrero, and M. Naveira (2005): "Are monetary policy reaction functions asymmetric? The role of non-linearity in the Phillips curve," *European Economic Review*, 49(2), 485–503.

DOLADO, J. J., R. M.-D. PEDRERO, AND F. J. RUGE-MURCIA (2004): "Nonlinear Monetary Policy Rules: Some New Evidence for the U.S.," *Studies in Nonlinear Dynamics & Econometrics*, 8(3), Article 2.

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

EL-GAMAL, M. A. (1993): "The Extraction of Information from Point Estimates," *Nonparametric Statistics*, 2, 369–378.

ELLISON, M. (2006): "The learning cost of interest rate reversals," *Journal of Monetary Economics*, 53(8), 1895–1907.

ELLISON, M., L. SARNO, AND J. VILMUNEN (2006): "Caution or Activism? Monetary Policy Strategies in an Open Economy," Unpublished manuscript.

ELLISON, M., AND N. VALLA (2001): "Learning, Uncertainty and Central Bank Activism in an Economy with Strategic Interactions," *Journal of Monetary Economics*, 48(1), 153–171.

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.

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.

HOLMQUIST, B. (1988): "Moments and Cumulants of the Multivariate Normal Distribution," *Stochastic Analysis and Applications*, 6(2), 273–278.

——— (1996): "Expectations of Products of Quadratic Forms in Normal Variables," *Stochastic Analysis and Applications*, 14(2), 149–164.

HONG, H., AND S. RADY (2002): "Strategic Trading and Learning about Liquidity," *Journal of Financial Markets*, 5, 419–450.

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

JUDGE, G. G., T.-C. LEE, AND R. C. HILL (1988): *Introduction to the Theory and Practice of Econometrics*. Wiley, New York, 2 edn.

JULIER, S. J., AND J. K. UHLMANN (1997): "A New Extension of the Kalman Filter to Nonlinear Systems," in *Proceedings of AeroSense: The 11th International Symposium on Aerospace/Defence Sensing, Simulation and Controls*, vol. MultiSensor Fusion, Tracking and Resource Management II, Orlando, Florida.

KENDALL, M. G., AND A. STUART (1963): *The Advanced Theory of Statistics*, vol. 1: Distribution Theory. Oxford University Press, London, 1 edn.

KENDRICK, D. A. (1978): "Non-convexities from probing an adaptive control problem," *Journal of Economic Letters*, 1(4), 347–351.

——— (1979): "Adaptive control of macroeconomic models with measurement error," in *Optimal Control of Econometric Models*, ed. by S. Holly, B. Rustem, and M. B. Zarrop, pp. 204–230. Macmillan, London.

——— (1982): "Caution and Probing in a Macroeconomic Model," *Journal of Economic Dynamics and Control*, 4(4), 149–170.

——— (2002): *Stochastic Control for Economic Models*. http://eco.utexas.edu/faculty/Kendrick, 2 edn.

KIM, D. H., D. R. OSBORNE, AND M. SENSIER (2005): "non-linearity in the Fed's monetary policy rule," *Journal of Applied Econometrics*, 20(5), 621–639.

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

Kotz, S., N. Balakrishnan, and N. L. Johnson (2000): *Continuous Multivariate Distributions*, vol. 1 of *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York, 2 edn.

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

MacRae, E. C. (1972): "Linear decision with experimentation," *Annals of Economic and Social Measurement*, 1, 437–447.

——— (1975): "An Adaptive Learning Role for Multi-Period Decision Problems," *Econometrica*, 43, 893–906.

Mizrach, B. (1991): "Non-convexities in a stochastic control problem with learning," *Journal of Economic Dynamics and Control*, 15, 515–538.

Moscarini, G., and L. Smith (2001): "The Optimal Level of Experimentation," *Econometrica*, 69(6), 1629–1644.

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.

Pearson, K., and A. W. Young (1918): "On the Product-Moments of Various Orders of the Normal Correlation Surface of two Variables," *Biometrika*, 12, 86–92.

Prescott, E. C. (1972): "The Multi-Period Control Problem Under Uncertainty," *Econometrica*, 40(6), 1043–1058.

Radner, R., and J. E. Stiglitz (1984): "The Non-Convexity in the Value of Information," in *Bayesian Models in Economic Theory: Studies in Bayesian Econometrics*, ed. by M. Boyer, and R. Kihlstrom, vol. 5. Elsevier, New York.

Rausser, G. C., and J. W. Freebairn (1974a): "Approximate Adaptive Control Solutions to U.S. Beef Trade Policy," *Annals of Economic and Social Measurement*, 3(1), 177–203.

——— (1974b): "Estimation of Policy Preference Function: An Application to U.S. Beef Import Policy," *The Review of Economics and Statistics*, 56(4), 437–449.

Rothschild, M. (1974): "Searching for Lowest Price When the Distribution of Prices is Uknown," *Journal of Political Economy*, 82(4), 689–711.

Tillmann, P. (2008): "Parameter Uncertainty and Non-Linear Monetary Policy Rules," Working paper, University of Bonn.

Trefler, D. (1993): "The Ignorant Monopolist: Optimal Learning with Endogenous Information," *International Economic Review*, 34(3), 565–581.

Triantafyllopoulos, K. (2002): "Moments and cumulants of the multivariate real and complex Gaussian distributions," Mimeo, Department of Mathematics, University of Bristol.

Tse, E., Y. Bar-Shalom, and L. Meier (1973): "Wide-Sense Adaptive Dual Control of Stochastic Nonlinear System," *IEEE Transactions on Automatic Control*, AC-18, 98–108.

Wieland, V. (2000): "Learning by Doing and the Value of Optimal Experimentation," *Journal of Economic Dynamics and Control*, 24(4), 501–535.

Winschel, V., and M. Krätzig (2008): "Solving, Estimating and Selecting Nonlinear Dynamic Models without the Curse of Dimensionality," SFB 649 Discussion Paper 2008-018, Humboldt-Universität zu Berlin, Berlin, Germany.

Yetman, J. (2000): "Probing Potential Output: Monetary Policy, Credibility and Optimal Learning under Uncertainty," Working Paper 2000-10, Bank of Canada.

Zampolli, F. (2005): "Optimal Monetary Policy in a Regime-Switching Economy: The Response to Abrupt Shifts in Exchange-Rate Dynamics," Working paper, Bank of England.

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