ELSEVIER Journal of Economic Dynamics & Control 26 (2002) 1359–1377

JOURNAL OF Economic Dynamics & Control

www.elsevier.com/locate/econbase

Learning and control in a changing economic environment☆

Günter W. Beck[a], Volker Wieland[a,b,*]

[a] *Goethe Universität Frankfurt, Frankfurt am Main, Germany*
[b] *European Central Bank, Frankfurt am Main, Germany*

Abstract

In this paper we investigate optimal Bayesian learning and control with lagged dependent variables and time-varying unknown parameters. We assess both the performance of alternative decision rules, as in the computationally-oriented dual control literature, as well as the dynamics of learning and convergence of beliefs, as in the more theoretically oriented Bayesian learning literature. Our numerical results indicate that the optimal decision rule involves a noticeable degree of experimentation for moderate to large levels of uncertainty. In most situations though, the optimal rule will remain less activist than a certainty-equivalent rule and induce gradualism. Exceptions occur when the process to be controlled is near the deterministic steady state. In this situation, we find that the decision-maker will repeatedly undertake costly experiments. The extent of optimal experimentation is lower if the unknown parameter is perceived to vary over time. In contrast to the fixed parameter case, however, parameter uncertainty is continually renewed and the incentive to experiment never ceases.

JEL classification: C44; C60; D81; D82

Keywords: Optimal control; Learning; Bayes rule; Parameter uncertainty; Time-varying parameters

1. Introduction

Optimal control with learning about unknown parameters has been applied to a variety of economic problems ranging from monopolistic pricing with unknown demand

☆ Volker Wieland served as a consultant to the Directorate General Research of the European Central Bank while preparing this paper. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting those of the European Central Bank. We are grateful for helpful comments from an anonymous referee of this journal and from seminar participants at Cambridge University, the Tinbergen Institute and the European Central Bank. Of course, the authors are responsible for any remaining errors.

* Corresponding author. Tel.: +49-69-798-25288; fax: +49-69-798-25272.
E-mail address: wieland@wiwi.uni-frankfurt.de (V. Wieland).

0165-1889/02/$ - see front matter
PII: S0165-1889(01)00075-6

to optimal investment with production uncertainty and even fiscal and monetary policy with imperfect knowledge about the macroeconomy. A well-known feature of such control problems is the possibility of a tradeoff between current control and estimation, that is, a tradeoff between maximizing expected current reward and experimenting to obtain more precise parameter estimates and thus improve expected future rewards.

During the past three decades two different literatures developed focusing on different aspects of this tradeoff. A computationally-oriented literature, which started originally in engineering, focused on comparing the performance of alternative decision rules in multi-period control problems. In particular, this research compared certainty-equivalent and cautionary decision-making with active probing. Early contributions to this line of research include Prescott (1972), MacRae (1972), Tse and Bar-Shalom (1973) and Bar-Shalom and Tse (1976). The methodology proposed by Bar-Shalom and Tse, often referred to as dual control, was generalized and applied to economic models by Kendrick (1978, 1979, 1981, 1982). Kendrick (1978) also discovered non-convexities, which imply multiple optima in the payoff function of the dual control problem. These non-convexities were investigated further by Mizrach (1991) and Amman and Kendrick (1994a, 1994b, 1995).

In parallel, a second literature with a more theoretical bent focused on learning and the asymptotic properties of the decision-maker's beliefs regarding the unknown parameters. Early contributions by Taylor (1974), Anderson and Taylor (1976) and Lai and Wei (1982) were considerably generalized in subsequent research on optimal Bayesian learning by Easley and Kiefer (1988), Kiefer and Nyarko (1989), Nyarko (1991) and Aghion et al. (1991). The second group of authors primarily focused on studying the possibility of incomplete learning and the conditions under which complete learning would occur in the long run. More recently, research within the Bayesian learning framework has also attempted to quantify the optimal extent of probing in specific economic models similar to the computationally-oriented dual control literature. Contributions include Kiefer (1989) and Wieland (1998, 2000a, 2000b). The numerical approach used in these papers is most closely related to Prescott (1972). Kiefer (1989) and Wieland (2000a and 2000b) use numerical approximations of the optimal decision rule (incorporating optimal probing) in order to study the likelihood of incomplete learning and the speed of convergence to the true parameter values in simple regression models. Wieland (1998) analyzes optimal experimentation in an economic model, which also includes a lagged dependent variable.

In this paper, we extend the analysis of Wieland (1998) to the case of time-varying unknown parameters. We focus on uncertainty regarding a parameter that is multiplicative to the decision variable, because this type of parameter is crucial for the tradeoff between current control and estimation. We compare certainty-equivalent, cautionary and optimal decision rules. Optimal decisions involve a certain degree of experimentation, which becomes apparent in a more aggressive response to past deviations of the dependent variable from the decision-maker's target value. Nevertheless, the optimal decision rule remains in most situations less aggressive (i.e. more gradualist) than a certainty-equivalent decision rule that completely disregards parameter uncertainty.

Once we allow for the possibility of time-variation in the unknown multiplicative parameter, we find that the optimal extent of experimentation is reduced. The reason is that the expected payoff to probing and obtaining a more precise parameter estimate is lower when the decision-maker expects the parameter to change again. However, time-variation in the unknown parameter also implies that the uncertainty regarding this parameter is renewed again and again. Thus, in contrast to the case of unknown fixed parameters the incentive to experiment never disappears in the case of unknown time-varying parameters. As a result, the decision-maker will be willing to undertake costly experiments repeatedly. In other words, the decision-maker will tolerate some level of steady-state fluctuations, because they provide information about the unknown time-varying parameters.

2. The decision problem

2.1. The environment

The decision-maker is faced with a generic linear stochastic process of the following form:

$$x_t = \alpha + \beta_t u_t + \gamma x_{t-1} + \varepsilon_t, \quad \text{where } \varepsilon_t \sim N(0, \sigma_\varepsilon). \tag{1}$$

In terms of notation, we follow the dual control literature. The state variable, that is, the dependent variable in the regression equation is denoted by x_t while the right-hand side control variable is denoted by u_t.[1] Finally, ε_t is a normally distributed noise term with zero mean and variance σ_ε.

The linear process in (1) contains three potentially unknown parameters that may have to be estimated by the decision-maker, the intercept α, the multiplicative slope parameter β and the persistence parameter γ on the lagged dependent variable. In the following analysis we restrict our attention to the case where α and γ are known to the decision-maker and only β is unknown and has to be estimated. The reason for focusing exclusively on incomplete information regarding β is that this parameter is multiplicative to the decision variable u_t and therefore central to the trade-off between current control and estimation. Intuitively, this becomes clear when one considers keeping u_t fixed at some value versus changing it around. For example, if u were kept fixed at zero for some time, regressing observations of x on lagged values of itself and u would not help in estimating β any more precisely. However, if u varied a lot, precise estimates of β could be obtained fairly quickly.

The parameter β in (1) has a time subscript indicating that it may vary. We consider the case of the random walk process

$$\beta_t = \beta_{t-1} + \eta_t, \quad \text{where } \eta_t \sim N(0, \sigma_\eta), \tag{2}$$

[1] Wieland (2000a, 2000b) instead followed the Bayesian learning literature (e.g., Kiefer and Nyarko, 1989) by using y_t to denote the dependent variable and x_t to denote the right-hand side decision variable in the regression equation.

where the normally distributed innovations η_t with zero mean and variance σ_η have a permanent effect on the unknown parameter β.[2]

2.2. Learning

The decision-maker's beliefs regarding β prior to choosing u_t and observing x_t are described by a conditional normal distribution with mean $b_{t|t-1}$ and variance $v^b_{t|t-1}$:

$$E_{t-1}[\beta_t] = b_{t|t-1} = b_{t-1},$$
$$VAR_{t|t-1}[\beta_t] = v^b_{t|t-1} = v^b_{t-1} + \sigma_\eta. \tag{3}$$

Following the choice of u_t and realization of the shock ε_t the new observation x_t becomes available and the decision-maker updates his beliefs according to

$$b_{t|t} = b_{t|t-1} + v^b_{t|t-1}(u_t)F^{-1}(x_t - \alpha - b_{t|t-1}u_t - \gamma x_{t-1}),$$
$$v^b_{t|t} = v^b_{t|t-1} - v^b_{t|t-1}(u_t)F^{-1}(u_t)v^b_{t|t-1}, \tag{4}$$

where $F = (u_t)v^b_{t|t-1}(u_t) + \sigma_\varepsilon$.

Here, learning is equivalent to Kalman filtering or recursive least squares with time-varying parameters.[3] With the assumption of normal priors and normal shocks it is also equivalent to Bayesian learning with conjugate prior. As can be seen from the second equation in (4), the updating equation for the variance of the parameter estimate is a deterministic process. $v^b_{t|t}$ is a function of u_t, of last period's variance $v^b_{t-1|t-1}$ and of the variance of the innovations to the unobserved parameter σ_η but not of the random shocks η and ε. The updating equation for the variance is non-increasing if σ_η is equal to zero.

As can be seen from the first equation in (4), updating of the mean of the policy-maker's beliefs that $b_{t|t}$ is a stochastic process, because $b_{t|t}$ depends on the forecast error, $x_t - (\alpha + b_{t|t-1}u_t + \gamma x_{t-1})$. This forecast of x_t is conditional on the decision-maker's prior beliefs regarding the parameters, the realization of x in the previous period and current the decision u_t. The magnitude of the update is decreasing in the conditional variance of x_t (denoted by F in (4)) and the variance of the shocks to the dependent variable σ_ε, but increasing in the variance of last period's point estimate v^b_{t-1} and the variance of the shocks to the unobserved parameter σ_η.

---

[2] An interesting alternative would be a specification that allows for mean reversion such as $\beta_t = \rho(\beta_{t-1} - \bar{\beta}) + \eta_t$ with $|\rho| < 1$. This specification would be of interest for applications in which the unknown parameter is unlikely to stray far from its long-run average. For our purposes here, this specification has the drawback that one either needs to treat ρ and $\bar{\beta}$ as known to the decision-maker (an assumption which is employed in a model with learning analyzed by Balvers and Cosimano (1994) but could be considered a bit artificial) or solve a substantially more complicated learning problem where these parameters are unobserved. We discuss a conjecture regarding the optimal extent of experimentation when β_t is mean-reverting in Section 3.

[3] For a derivation of the updating equations see Harvey (1992).

2.3. Optimal control

We endow the decision-maker with a quadratic per-period expected loss function

$$L(x_t, u_t) = E_{t-1}[(x_t - x^*)^2 + \omega(u_t - u^*)^2], \tag{5}$$

which is standard in the dual control literature with target values for the state and control variables, x^* and u^*, respectively, and a weight of ω on the deviations of the control variable from target.

The dynamic optimization problem with infinite horizon and discounting is then given by

$$\underset{[u_t]_{t=0}^{\infty}}{\text{Min}} \quad \text{E}\left[\sum_{t=0}^{\infty} \delta^t((x_t - x^*)^2 + \omega(u_t - u^*)^2) \,|\, (x_0, u_0, b_0, v_0^b)\right]$$

$$\text{s.t.} \quad \text{Eqs. (1) and (4)}, \tag{6}$$

where δ is the discount factor. The dynamic problem has three state variables, x, b and v^b. Although the stochastic process to be controlled is linear and the loss function quadratic, the updating equations for the beliefs are non-linear. Thus, this dynamic optimization problem falls outside the class of linear-quadratic problems. The associated Bellman equation is given by

$$\begin{aligned} &V(x_{t-1}, b_{t-1}, v_{t-1}^b) \\ &= \underset{u_t}{\text{Min}} \left[L(x_{t-1}, b_{t-1}, v_{t-1}^b, u_t) + \delta \int V(x_t, b_t(x_{t-1}, b_{t-1}, v_{t-1}^b, x_t, u_t), v_t^b(v_{t-1}^b, u_t)) \right. \\ &\qquad \left. \times f(x_t | x_{t-1}, b_{t-1}, v_{t-1}^b, u_t)\, \mathrm{d}x \right], \end{aligned} \tag{7}$$

where $b_t(\cdot)$ and $v_t^b(\cdot)$ stand for the updating equations in (4) and $f(x_t|\cdot)$ denotes the distribution of x_t conditional on the decision-maker's beliefs and action. Given the preceding assumptions $f(x_t|\cdot)$ is a normal distribution with mean $\alpha + b_{t|t-1}u_t + \gamma x_{t-1}$ and variance F. Following Easley and Kiefer (1988) and Kiefer and Nyarko (1989), it can be shown that a stationary optimal policy exists and the value function is continuous and satisfies the above Bellman equation. The stationary optimal policy, which we denote by $H(\cdot)$, is a function of the three state variables $(x_{t-1}, b_{t-1}, v_{t-1}^b)$:

$$u_t = H(x_{t-1} - x^*, b_{t-1}, v_{t-1}^b, \alpha, \gamma, \sigma_\varepsilon, \sigma_\eta, \delta, u^*). \tag{8}$$

It is also influenced by parameters affecting the environment and the decision-maker's preferences such as the variances of the two noise terms, the degree of persistence in the linear process or the discount factor.

Policy and value functions can be obtained using an iterative algorithm based on the Bellman equation starting with an initial guess of the value function $V(\cdot)$. However, the integration in (7) cannot be carried out analytically since the updating functions for the beliefs are non-linear functions of x and u. Instead, we use numerical dynamic

programming methods, which will be described later, to obtain approximations to the optimal policy and value functions.

3. Comparing alternative decision rules

As suggested in the earlier dual control literature we compare three different decision rules, that is, the certainty-equivalent, cautionary and dynamically optimal rules. With regard to the loss function we restrict attention to the case of $\omega = 0$, for which the certainty-equivalent rule implies full stabilization (i.e. the absence of gradualism).

3.1. The certainty-equivalent rule

The certainty-equivalent rule corresponds to the optimal strategy if one disregards parameter uncertainty and goes ahead with the best available parameter estimates. While this certainty-equivalent approach is well-known to be optimal in standard linear quadratic problems with measurement error, it is definitely not optimal in the case of multiplicative parameter uncertainty. Nevertheless, it constitutes a useful benchmark. For $\omega = 0$ we obtain

$$u_t = -\left(\frac{\gamma}{b_{t-1}}\right)(x_{t-1} - x^*) - \left(\frac{\alpha}{b_{t-1}}\right). \tag{9}$$

The certainty-equivalent optimal value of u at time t deviates from the estimated neutral value (or rest point) of $-\alpha/b_{t-1}$, whenever the lagged state variable x_{t-1} has deviated from the target value of x^*. The difference from the estimated neutral value is intended to fully offset the predictable impact of last period's state on the current state, which in turn depends on the degree of persistence γ. Of course, the smaller the multiplicative parameter β, the greater needs to be the offsetting move in the control variable.

The certainty-equivalent rule results in complete stabilization of predictable fluctuations in x. Observable fluctuations in x will only be due to the initial impact of unpredictable shocks and forecast errors. Thus, the decision rule does not exhibit *gradualism*, which would mean that predictable movements in x are only partially offset resulting in gradual adjustment paths.[4] One potential source of such gradualism would be a preference parameter $\omega > 0$, which assigns some weight to variations of the control variable in the loss function.[5] In the following analysis, we explore instead to what extent parameter uncertainty can be a source of gradualism.

3.2. The cautionary myopic rule

The second decision rule we consider is one which takes into account the degree of uncertainty regarding the multiplicative parameter β but disregards the dynamic link

[4] For a discussion of gradualism in the context of monetary policy and the Phillips curve and simulation results the reader is referred to Wieland (1998).

[5] In this case, the dynamically optimal decision rule under certainty (and as a consequence the certainty-equivalent rule) would also depend explicitly on the discount factor δ.

between current and future beliefs that arises from learning. This rule is myopic and ignores the incentive for experimentation. It is obtained by minimizing current expected loss conditional on the decision-maker's beliefs regarding the unknown parameter and corresponds to

$$u_t = -\left(\frac{\gamma(b_{t-1})}{b_{t-1}^2 + v_{t-1}^b + \sigma_\eta}\right)(x_{t-1} - x^*) - \left(\frac{\alpha(b_{t-1})}{b_{t-1}^2 + v_{t-1}^b + \sigma_\eta}\right). \tag{10}$$

The response to deviations of x_{t-1} from x^* is equal to the certainty-equivalent response whenever the variance of the parameter estimate ($v_{t|t-1}^b = v_{t-1}^b + \sigma_\eta$) is equal to zero, but smaller when there is uncertainty and the variance positive. Compared to the certainty-equivalent rule this cautionary, less aggressive response results in gradualism. The decision-maker will take several periods to offset the predictable impact of a shock (i.e. the impact on future values of x as a result of the initial effect of the shock on the current value of x) rather than offsetting it completely one period after the initial shock occurred. The estimated rest point is also biased towards zero, since $(\alpha(b_{t-1}))/(b_{t-1}^2 + v_{t-1}^b + \sigma_\eta) < \alpha/(b_{t-1})$ whenever β is uncertain.

In the following, we explore to what extent the *dynamically optimal decision rule* deviates from the other two rules. The extent of probing or experimentation is best defined as the difference between the cautionary and dynamically optimal rule. The extent of gradualism, however, is best measured with respect to the certainty-equivalent rule. In particular, if the optimal rule is more activist in terms of the response to the lagged dependent variable than the cautionary rule but remains less activist than the certainty-equivalent rule it could still be said to induce gradualism.

3.3. The numerical algorithm for computing the optimal rule

As noted above, the dynamically optimal rule in the presence of learning about an unobserved multiplicative parameter cannot be derived analytically. Instead, we approximate the optimal rule numerically. Our numerical dynamic programming algorithm iterates over the Bellman Eq. (7) by means of value and policy iterations until it converges within a narrow range of the true value function. A value iteration contains the following steps:

1. It begins with a guess of the value function $V(\cdot)$. In the first iteration, the loss function resulting from the myopic rule can be used as initial guess.
2. An update of the value function is computed on a grid over the state variables by solving the optimization problem on the right-hand side of the Bellman equation. As there are three continuous state variables, the dimension of this DP problem is quite large. The grid-based approximation of the true value function is saved and used in the subsequent iteration. The following set of operations has to be carried out for every grid point.
3. Integration step: since the dependent variable x_t is normally distributed this step uses Gauss–Hermite quadrature to approximate the expectation of next period's value function (the continuation value). To evaluate this expectation for different possible

outcomes of x_t resulting in beliefs that do not fall exactly on grid points, one needs to interpolate the last iteration's guess of the value function. [6]

4. Optimization step: the optimization problem is characterized by multiple optima. Thus, a fairly robust search routine is needed. Fortunately though, the optimization is one dimensional. We start with a rough grid-search and then proceed with golden section search.

Convergence of value iterations is guaranteed by Blackwell's sufficiency conditions of monotonicity and discounting. The number of iterations needed can be reduced by conducting policy iterations after each value iteration. For a more detailed mathematical exposition of this numerical DP algorithm, see the appendix of Wieland (2000a).

As mentioned in the introduction, dual control according to Bar-Shalom and Tse (1976) and Kendrick (1981) provides a numerical approach for solving the same type of learning problem as considered in this paper. There are some important differences in methodology though:

- While the dual control approach typically involves a first- or second-order linear approximation, the DP algorithm used here directly takes into account the non-linearity of the updating equations that is the crucial feature of the learning problem.
- Contrary to the dual control approach, which approximates ex-post payoffs for a given initial belief about the unknown parameters for alternative draws of shocks (Monte Carlo simulations), the DP algorithm used here approximates ex-ante payoffs and policies for a large range of initial beliefs and any sequence of shocks.

For these reasons, we expect the DP algorithm to provide more insight into the implications of the non-linearity introduced by learning and a more precise numerical approximation. However, the application of the DP algorithm is limited to cases with few state variables. For example, if one uses 100 grid points in each dimension, the number of computations increases 100-fold with each additional state variable. The dual control approach, which relies on Monte Carlo simulations instead can also be applied to larger-scale problems. In this sense, the two approaches are complementary.

3.4. Numerical results

Fig. 1 provides a first comparison of the certainty-equivalent (solid line), cautionary (dashed line) and dynamically optimal (solid line with dots) decision rules for a given belief regarding the unknown parameter β. This belief corresponds to a normal distribution with mean $b = 0.5$ and a standard deviation $\sqrt{v^b} = -0.5$. Thus, the initial estimate is very imprecise. The horizontal axis in Fig. 1 measures the deviation of the state x_{t-1} from the decision-maker's target x^*. The vertical axis denotes the deviation of the control variable u_t from its rest point. [7]

[6] Except in the first iteration when the one-period expected loss is used as initial guess. We use bilinear interpolation.

[7] Note, since α is known and set equal to zero in this case, the neutral value of u is also equal to zero for any estimate of β.

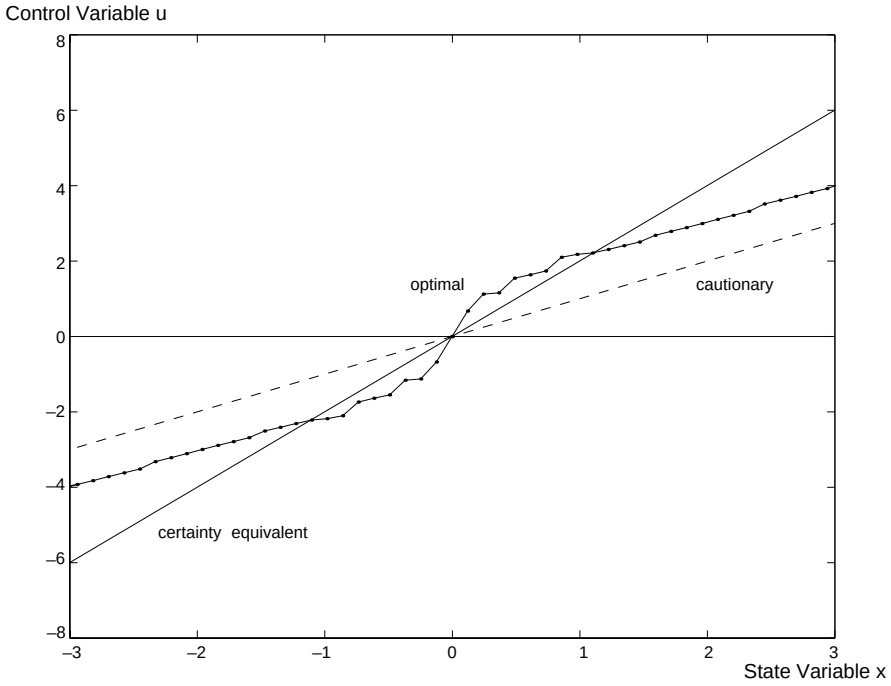

Fig. 1. Three alternative decision rules—certainty-equivalent, cautionary and optimal. Initial beliefs are set to $(b_0 = -0.5, v_0^b = 0.25)$. The parameter settings are $(\alpha = 0, \gamma = 1, \sigma_\varepsilon = 1, \sigma_\eta = 0, \omega = 0, u^* = 0, \delta = 0.95)$.

Both the certainty-equivalent and the cautionary rule respond linearly to x_{t-1}. The slope of these rules is determined by the response coefficients in (9) and (10), respectively. Since the estimate of β is uncertain, the cautionary rule responds less aggressively and does not completely offset the impact of the past deviation from target on the current state. As a result, an unexpected shock ε will have a persistent effect on x, which will only disappear gradually over time.

Our numerical results indicate that the dynamically optimal rule always responds more aggressively than the cautionary rule. Thus, it incorporates a significant degree of probing or experimentation.[8] Except for values of x near target, the degree of experimentation is fairly constant, i.e. not proportional to the size of the past deviation

[8] To our knowledge, no theoretical results regarding the form of the dynamically optimal decision rule in this framework are available from the literature. However, our numerical findings concerning the increased aggressiveness of the optimal decision rule relative to the cautionary rule are consistent with theoretical findings of Prescott (1972) in a simpler framework without lagged dependent variable. For a learning problem with (γ, α, ω) all equal to zero and with β constant, Prescott (1972) proves that the optimal decision will be larger in absolute value than the one which minimizes expected loss in the current period, i.e. more aggressive than cautionary myopic rule. Even in this simpler model proving this property of the optimal decision rule is rather involved and a quantitative assessment of the optimal extent of experimentation still needs to rely on numerical analysis.

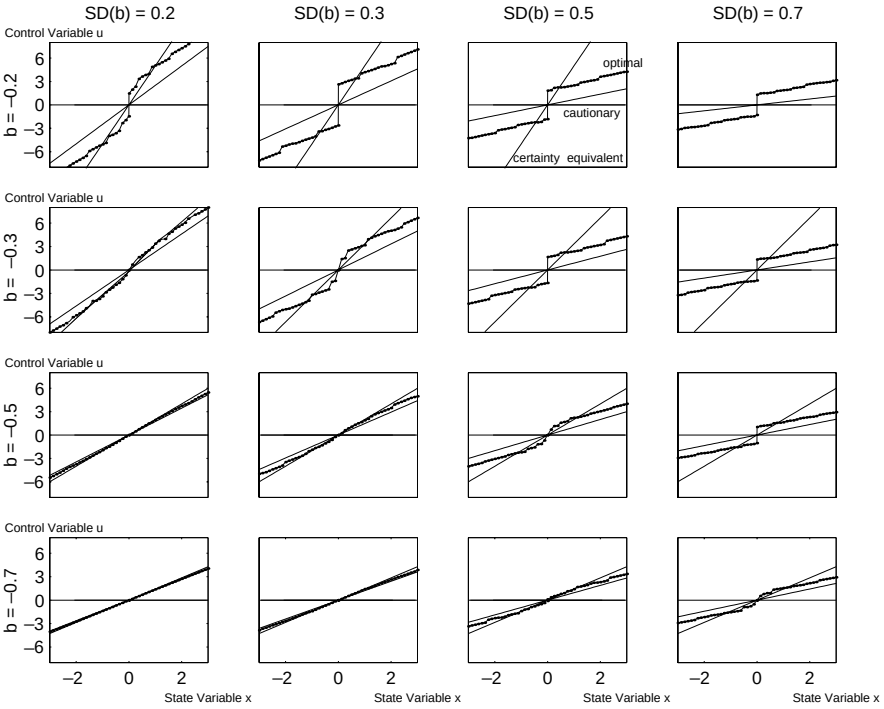

Fig. 2. A comparison for a wide range of beliefs. The parameter settings are ($\alpha = 0, \gamma = 1, \sigma_\varepsilon = 1, \sigma_\eta = 0, \omega = 0, u^* = 0, \delta = 0.95$).

from target. Furthermore, for moderate to large deviations of x from target the optimal rule remains less aggressive than the certainty-equivalent rule. In other words, even the dynamically optimal rule with experimentation implies a gradualist response to moderate and large deviations.

When x is near target, however, the dynamically optimal response exceeds the certainty-equivalent response. For this range, the incentive to experiment is strong enough to result in activist rather than gradualist decision-making.

Of course, the comparison shown in Fig. 1 depends on the specific set of beliefs. Fig. 2 provides a set of 16 panels, which indicate that these findings extend to a much wider range of possible beliefs (combinations of b and v^b). As we move horizontally from left to right, each panel in Fig. 2 is associated with a higher standard deviation of the estimate, $\sqrt{v^b} \in \{0.2, 0.3, 0.5, 0.7\}$. Similarly, as we move vertically from top to bottom, each panel is associated with an estimate b that increases in absolute value, $b \in \{-0.2, -0.3, -0.5, -0.7\}$. The third panel in the third row of Fig. 2 replicates Fig. 1.

Over this wider range of beliefs, we confirm that the dynamic rule tends to respond more aggressively to past deviations from target than the cautionary rule. Thus, it incorporates experimentation. However, the degree of experimentation is fairly small when the estimate is quite precise (lower-left corner panel). We also confirm that the

degree of experimentation is fairly constant for medium to large deviations of x. Over this range of deviations the optimal decision rule also remains less aggressive than the certainty-equivalent rule. Thus, it still implies gradualism.

Exceptions occur when x is near target. Then the optimal rule is more activist in order to obtain better estimates of the unknown parameter. This effect is particularly pronounced for very high degrees of uncertainty (upper right corner panel). In this situation, the optimal decision rule exhibits a clear discontinuity at zero. This discontinuity indicates that the two local optima (associated with a positive and negative control value) are exactly equal. These multiple optima arise from the non-convexity that was also encountered by Kendrick and others in the dual control literature.

The implications of this discontinuity are quite interesting. It implies that in the deterministic steady state, where it is optimal from a current-expected-loss perspective to set the control to its rest point, the dynamically optimal decision actively induces perturbations purely in order to obtain more precise parameter estimates. The motivation for this behavior is the following. The decision-maker knows that shocks will occur in the future, which will need to be counteracted. Future stabilization, however, will be much more effective with more precise estimates of the unknown parameter. Thus, it is optimal to induce somewhat more variation in states where x is close to target in order to learn more about the parameters that are crucial for controlling large deviations of x from target in the future.

Having investigated the optimal extent of experimentation as a function of the beliefs (b, v^b) and the state x, it is of interest to assess how it varies with certain parameters of the stochastic process such as the variance of the shocks σ_ε and the degree of persistence γ or with preference parameters such as the discount factor δ.

Fig. 3 shows that the optimal extent of experimentation increases with the variance of shocks σ_ε for a wide range of possible beliefs. Each panel compares the *difference* between optimal and cautionary decision rules (i.e. the extent of experimentation) for two scenarios. In the first case (solid line) the variance of shocks σ_ε is equal to 0.5, in the second case (dotted line) it is equal to 1. As is apparent from each panel the degree of experimentation is larger in the second case. Unfortunately, the numerical precision of this difference is not great. This is not surprising if one recalls that the algorithm approximates the value function $V(\cdot)$. The policy function $H(\cdot)$ is related to the first derivative of the value function, and therefore already approximated with less precision. The extent of experimentation as a function of parameters such as σ_ε essentially means differentiating the policy function, which further increases the imprecision of the numerical approximation.

Fig. 4 serves to show that the optimal extent of experimentation increases with the degree of persistence. The persistence parameter γ is set to 0.5 (solid line) and 1 (dotted line), respectively. Fig. 5 similarly shows that the decision-maker is more willing to experiment the more he cares about future rewards, that is, the greater the discount factor. The comparison is for $\delta = 0.75$ and $\delta = 0.95$, respectively.

Of more fundamental interest may be the question whether the incentive to experiment is greater in an environment with *constant* or with *time-varying* unknown parameters. The answer to this question is provided in Fig. 6. The optimal extent of experimentation when β is a constant parameter ($\sigma_\eta = 0$) is shown by the dotted line.

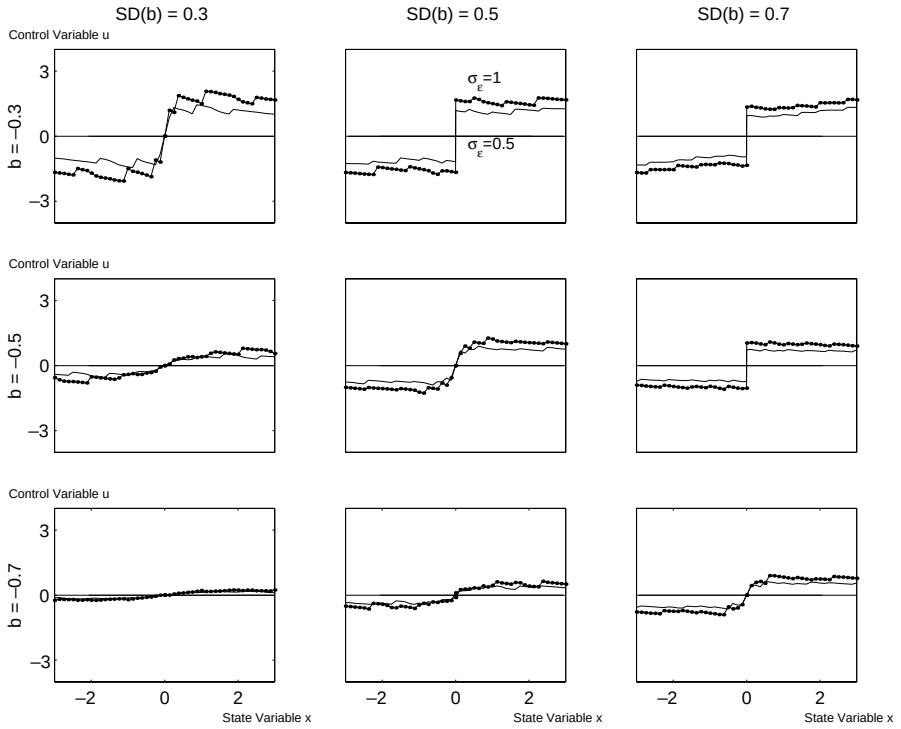

Fig. 3. Experimentation increases with the variance of shocks. The parameter settings are ($\alpha = 0, \gamma = 1$, $\sigma_\eta = 0, \omega = 0, u^* = 0, \delta = 0.95$). Each panel plots the difference between optimal and cautionary policy for two different values of the variance of σ_ε ($\sigma_\varepsilon = 1$ and $\sigma_\varepsilon = 0.5$).

The case of a time-varying, random-walk parameter, where any shock η has a permanent impact on the unknown parameter β is depicted by the solid line ($\sigma_\eta = 0.04$). In spite of the difficulty in providing a very precise approximation to the extent of experimentation, it is clear from this comparison that the degree of experimentation is smaller when the parameter β is believed to change over time. The reason of course, is that the information gain resulting from experimentation is shorter-lived when the unknown parameter changes over time and follows a random walk. [9]

4. Convergence

Having compared the optimal decision rule with learning to two alternative rules of interest, we now turn to the question of convergence of beliefs and actions. We consider three different parameterizations of the linear process defined by (1).

[9] From this result, we conjecture that experimentation may be relatively more beneficial if the unobserved parameter is assumed to exhibit mean revision, as in the example given in footnote 2. Although β also changes in this case, the information gain from experimentation may be more valuable since β remains nearer to some average value. Whether this conjecture holds is a question for future study.

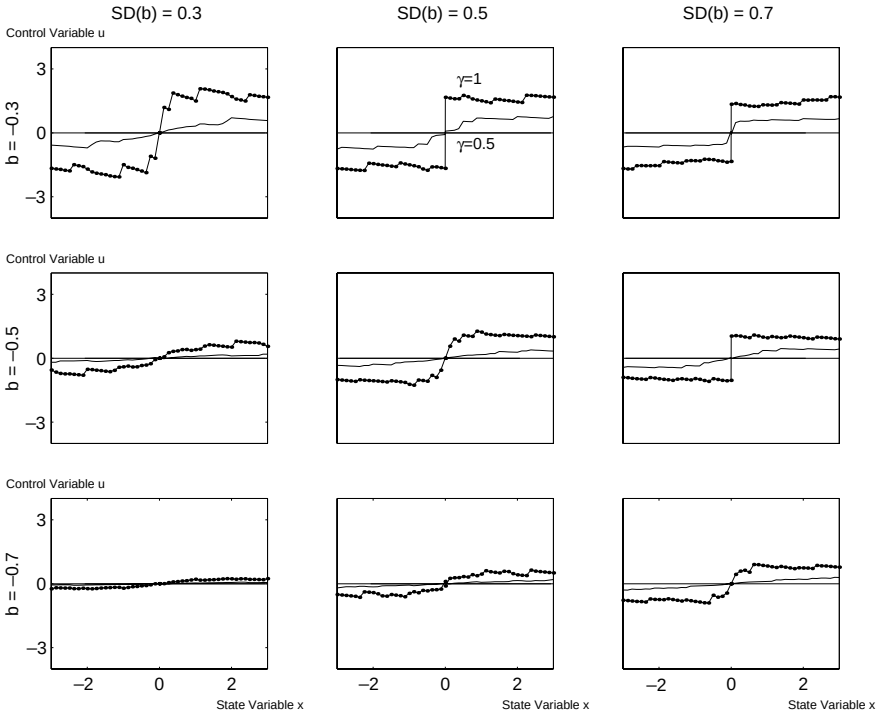

Fig. 4. Experimentation increases with the degree of persistence. The parameter settings are ($\alpha = 0, \sigma_\varepsilon = 1$, $\sigma_\eta = 0, \omega = 0, u^* = 0, \delta = 0.95$). Each panel plots the difference between optimal and cautionary policy for two different values of the variance of γ ($\gamma = 1$ and $\gamma = 0.5$).

(*A*) *Simple regression with constant parameters* ($\gamma = 0$, β *constant*)

$$x_t = \alpha + \beta u_t + \varepsilon_t. \tag{11}$$

In this scenario, the value chosen for u, if the parameters α and β were known, is constant. Kiefer and Nyarko (1989) show that beliefs (b, v_b) and actions u converge in the limit. However, if u converges "too" quickly, then beliefs may converge to incorrect values and the limit action may be incorrect (at least when both, α and β, are unknown). For a further characterization of potential limit beliefs, see Kiefer and Nyarko (1989). Wieland (2000a, 2000b) computes optimal policies for such regressions and studies the likelihood of incomplete learning and the properties of time series generated under alternative policies.

(*B*) *Regression with lagged dependent variable and constant parameters* ($\gamma \neq 0$, β *constant*)

$$x_t = \alpha + \beta u_t + \gamma x_{t-1} + \varepsilon_t. \tag{12}$$

In this case, the random shocks ε have a persistent effect on the dependent variable x. There is need for repeated stabilizing action by the decision-maker and the right-hand

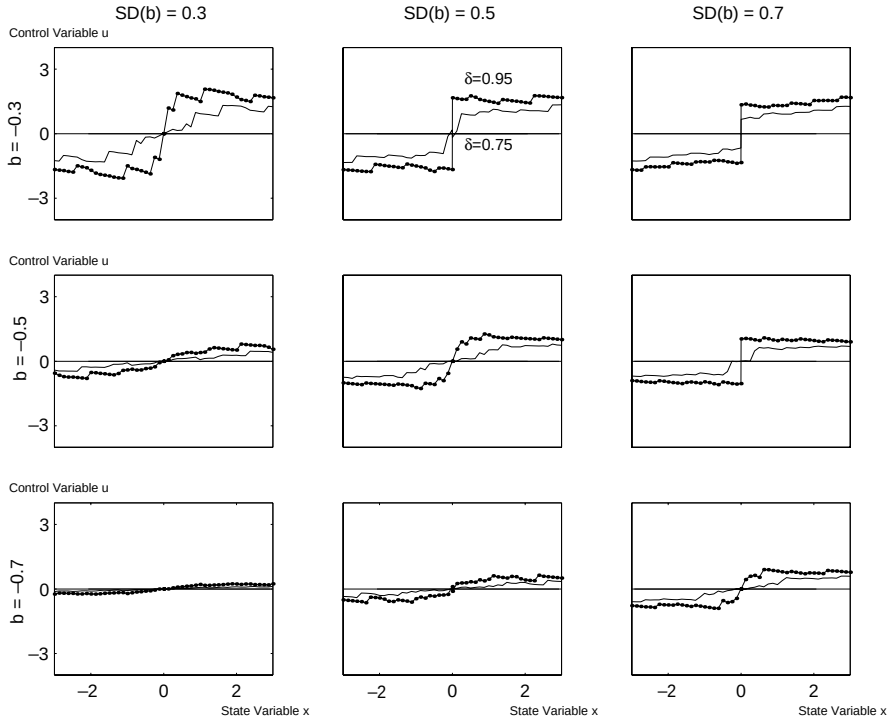

Fig. 5. Experimentation increases with the discount factor. The parameter settings are ($\alpha = 0, \gamma = 1$, $\sigma_\varepsilon = 1, \sigma_\eta = 0, \omega = 0, u^* = 0$). Each panel plots the difference between optimal and cautionary policy for two different values of the variance of δ ($\delta = 0.95$ and $\delta = 0.75$).

side variable u never settles down. For parameterization (A), Kiefer and Nyarko (1989) have shown that beliefs would converge with probability 1 to the truth if u does not converge. This finding implies convergence to the truth in parameterization (B).[10] Since beliefs converge to the truth, decision making converges to the optimal rule under certainty. Thus, one can derive the stochastic steady state of the control variable u from (9) by replacing b with β.

With constant parameters the incentive to experiment is temporary. It disappears over time as parameter estimates become more precise. This effect is shown in Fig. 7. It depicts a simulation of the paths of the state x, the control u and the beliefs (b, v^b) with all future shocks ε set to zero. The simulation starts from the initial belief $(b_0 = -0.5, v_0^b = 0.25)$, the same as in Fig. 1, with the true value of β set equal to its expectation. This belief is characterized by a high degree of uncertainty.

The initial state x_0 is set *close but not equal* to the target x^*. Thus, we start in the range where the optimal policy is more aggressive than the certainty-equivalent policy and some additional perturbation of the system in order to obtain more precise

[10] See also the discussion in this respect in Wieland (1998).

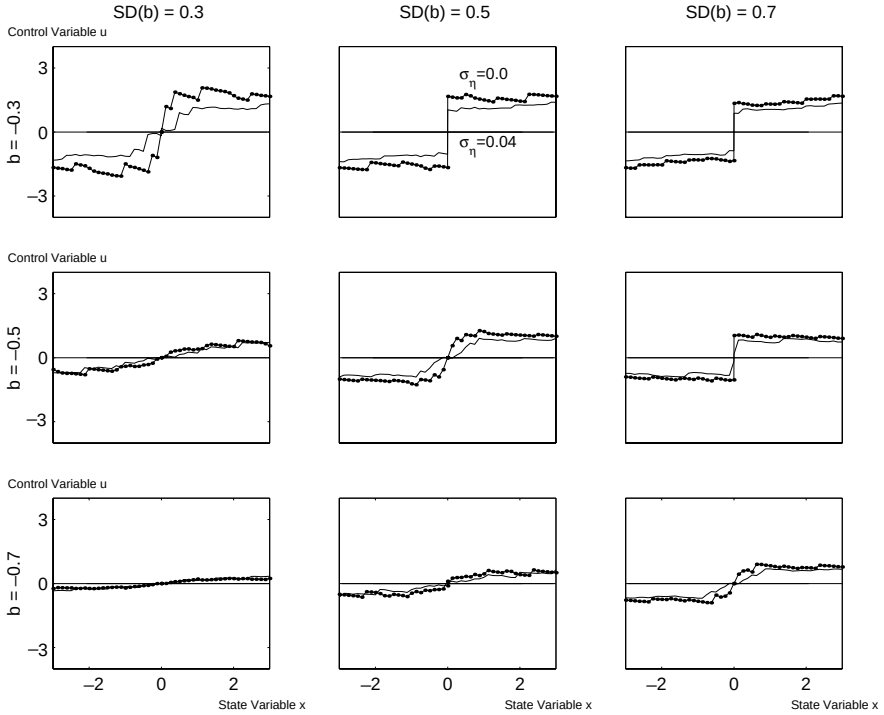

Fig. 6. Experimentation is lower with time-varying parameter. The parameter settings are ($\alpha = 0, \gamma = 1$, $\sigma_\varepsilon = 1, \omega = 0, u^* = 0, \delta = 0.95$). Each panel plots the difference between optimal and cautionary policy for two different values of the variance of σ_η ($\sigma_\eta = 0$ and $\sigma_\eta = 0.04$).

parameter estimates in the future is optimal. This can be seen from two top panels of Fig. 7. The top left panel indicates that the small initial deviation from target is positive (0.01) and consequently induces a positive response in the control variable (top right panel). The paths of x and u simulated in these two panels indicate that the optimal decision rule perturbs the system and induces fluctuations in x for about 20 periods. It is important to understand that none of these fluctuations are caused by shocks, since current and future shocks have been set to zero. Rather, the fluctuations in x are generated by the decision-maker in order to improve the precision of future estimates of β.

The gain in the precision of the parameter estimates is directly apparent from the decline in the variance, v^b, shown in the lower right panel of Fig. 7. Once this variance has fallen far enough, in this case to about 0.1 given a point estimate of -0.5, fluctuations in the control and state variables essentially cease, because the incentive to experiment has disappeared.

The lower right panel of Fig. 7 reports the associated path of the point estimate b_t, which remains constant throughout the whole period. This should not be surprising, because the parameter estimate is only revised in response to forecast errors as

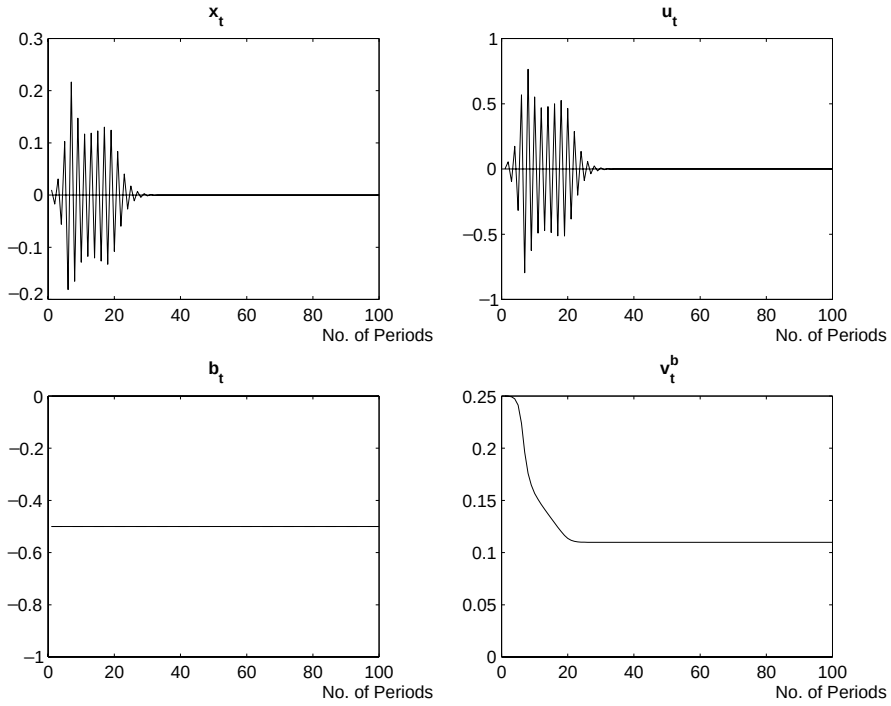

Fig. 7. Dynamic simulation with constant parameters—no shocks. Initial beliefs are set to ($b_0 = -0.5$, $v_0^b = 0.25$). The initial state is $x_0 = 0.01$. The parameter settings are ($\alpha = 0, \gamma = 1, \sigma_\varepsilon = 1, \sigma_\eta = 0$, $\omega = 0, u^* = 0, \delta = 0.95$). The shocks ε are set to zero and $\beta = b_0$.

discussed in Section 2. As long as no unexpected shocks occur, the forecasts of the decision-maker are correct and the initial parameter estimate is confirmed.

(*C*) *Regression with lagged dependent variable and time-varying parameter* ($\sigma_\eta > 0$)

$$x_t = \alpha + \beta_t u_t + \gamma x_{t-1} + \varepsilon_t. \tag{13}$$

As in parameterization (B), actions u do not converge, because the shocks ε continuously excite the system. Now however, a long-run limit for the beliefs (b, v^b) does not exist. Since the decision-maker assumes that the unknown parameter follows a random walk, the variance of the parameter estimate increases every period by the variance of the shock to the unobserved parameter, that is σ_η. As a result, uncertainty and consequently the incentive to experiment remain high, and the decision-maker continues to induce some level of fluctuations. This effect is illustrated in Fig. 8, which depicts a simulation starting from the same initial conditions as the simulation with a constant unobserved parameter in Fig. 7. Again, all current and future shocks (both ε and η) are set to zero. As can be seen from the top two panels, the decision-maker again starts to induce fluctuations in order to improve the precision of the parameter estimate.

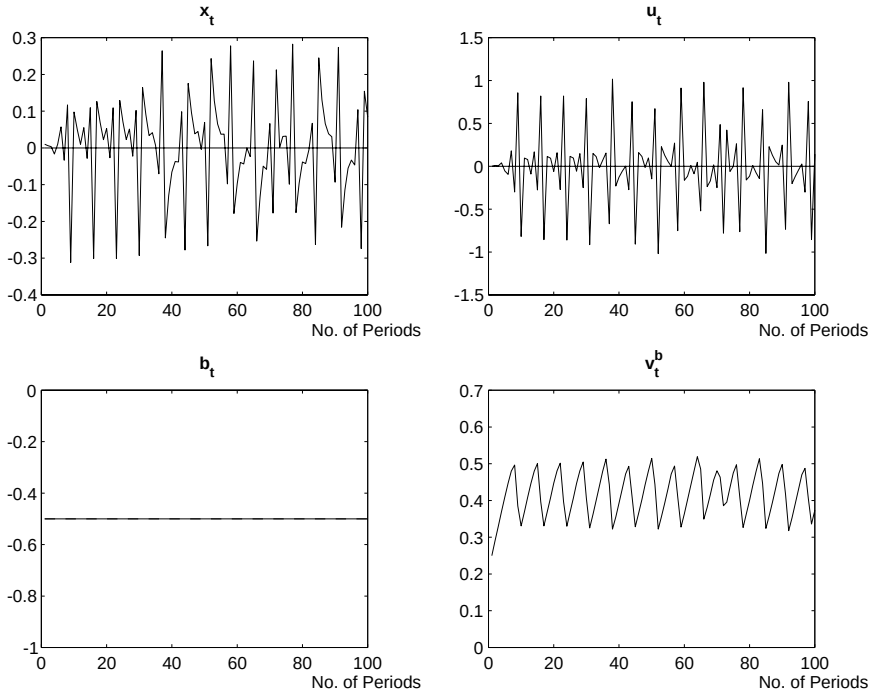

Fig. 8. Dynamic simulation with time-varying parameter—no shocks. Initial beliefs are set to ($b_0 = -0.5$, $v_0^b = 0.25$). The initial state is $x_0 = 0.01$. The parameter settings are ($\alpha = 0, \gamma = 1, \sigma_\varepsilon = 1, \sigma_\eta = 0.1$, $\omega = 0, u^* = 0, \delta = 0.95$). The shocks ε and η are set to zero and $\beta = b_0$.

However, because uncertainty regarding this parameter is renewed in every period (lower right panel), the incentive to experiment never ceases and the system exhibits a small level of endogenous fluctuations even when all stochastic shocks are set to zero.

5. Conclusion

Using numerical approximations of optimal decision rules in a Bayesian learning problem with unknown, potentially time-varying parameters and lagged dependent variables, we found that the optimal behavior involves a noticeable degree of experimentation for moderate to large levels of parameter uncertainty. Nevertheless, the optimal decision rule, in most cases, remains less aggressive than a certainty-equivalent rule. Thus, even the dynamically optimal decision rule with experimentation tends to induce gradualism. Exceptions occur when the linear process to be controlled is near the deterministic steady state. In this case, the incentive to experiment is strong enough to induce activism, that is, a more aggressive action than under the certainty-equivalent rule. If the unknown multiplicative parameter is perceived to vary over time, the incentive to

experiment is reduced. However, since parameter uncertainty in the time-varying parameter case is renewed again and again, the incentive to experiment never disappears as would be the case with constant parameters. Possible applications of the framework studied in this paper arise, for example, in the areas of monopolistic pricing, investment and growth and optimal policymaking under uncertainty.

References

Aghion, P., Bolton, P., Harris, C., Jullien, B., 1991. Optimal learning by experimentation. Review of Economic Studies 58, 621–654.

Amman, H., Kendrick, D., 1994a. Active learning: Monte Carlo results. Journal of Economic Dynamics and Control 18 (1), 119–124.

Amman, H., Kendrick, D., 1994b. Nonconvexities in stochastic control models: an analysis. In: Cooper, W.W., Whinston, A.B. (Eds.), New Directions in Computational Economics. Kluwer Academic Publishers, Dordrecht, pp. 57–94.

Amman, H., Kendrick, D., 1995. Nonconvexities in stochastic control models. International Economic Review 36 (2), 455–475.

Anderson, T., Taylor, J.B., 1976. Some experimental results on the statistical properties of least squares estimates in control problems. Econometrica 44, 1289–1302.

Balvers, R., Cosimano, T., 1994. Inflation variability and gradualist monetary policy. Review of Economic Studies 61, 721–738.

Bar-Shalom, Y., Tse, E., 1976. Caution, probing and the value of information in the control of uncertain systems. Annals of Economic and Social Measurement 4 (2), 239–252.

Easley, D., Kiefer, N.M., 1988. Controlling a stochastic process with unknown parameters. Econometrica 56, 1045–1064.

Harvey, A., 1992. Time Series Models, 2nd Edition. MIT Press, Cambridge.

Kendrick, D., 1978. Non-convexities from probing an adaptive control problem. Economic Letters 1, 347–351.

Kendrick, D., 1979. Adaptive control of macroeconomic models with measurement error. In: Holly, S., Rustem, B., Zarrop, M.B. (Eds.), Optimal Control of Econometric Models. Macmillan, London, pp. 204–230.

Kendrick, D., 1981. Stochastic control for economic models. In: Economic Handbook Series. McGraw-Hill, New York.

Kendrick, D., 1982. Caution and probing in a macroeconomic model. Journal of Economic Dynamics and Control 4, 149–170.

Kiefer, N., 1989. A value function arising in the economics of information. Journal of Economic Dynamics and Control 13, 201–223.

Kiefer, N., Nyarko, Y., 1989. Optimal control of an unknown linear process with learning. International Economic Review 30, 571–586.

Lai, T., Wei, C., 1982. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. The Annals of Statistics 10 (1), 154–166.

MacRae, E., 1972. Linear decision with experimentation. Annals of Economic and Social Measurement 1, 437–447.

Mizrach, B., 1991. Nonconvexities in a stochastic control problem with learning. Journal of Economic Dynamics and Control 15, 515–538.

Nyarko, Y., 1991. The number of equations versus the number of unknowns: the convergence of Bayesian posterior processes. Journal of Economic Dynamics and Control 15, 687–713.

Prescott, E., 1972. The multi-period control problem under uncertainty. Econometrica 40 (6), 1043–1058.

Taylor, J.B., 1974. Asymptotic properties of multi-period control rules in the linear regression model. International Economic Review 15, 472–484.

Tse, E., Bar-Shalom, Y., 1973. An actively adaptive control for linear systems with random parameters. IEEE Transactions on Automatic Control AC-18, 109–117.

Wieland, V., 1998. Monetary policy under uncertainty about the natural unemployment rate. Finance and Economics Discussion Series, Board of Governors of the Federal Reserve System, pp. 98–122.
Wieland, V., 2000a. Learning by doing and the value of optimal experimentation. Journal of Economic Dynamics and Control 24, 501–534.
Wieland, V., 2000b. Monetary policy, parameter uncertainty and optimal learning. Journal of Monetary Economics 46, 199–228.