How Active is Active Learning: Value Function Method Versus an Approximation Method

Hans M. Amman[1] · Marco P. Tucci[2]

Abstract

In a previous paper Amman et al. (Macroecon Dyn, 2018) compare the two dominant approaches for solving models with optimal experimentation (also called active learning), i.e. the value function and the approximation method. By using the same model and dataset as in Beck and Wieland (J Econ Dyn Control 26:1359–1377, 2002), they find that the approximation method produces solutions close to those generated by the value function approach and identify some elements of the model specifications which affect the difference between the two solutions. They conclude that differences are small when the effects of learning are limited. However the dataset used in the experiment describes a situation where the controller is dealing with a nonstationary process and there is no penalty on the control. The goal of this paper is to see if their conclusions hold in the more commonly studied case of a controller facing a stationary process and a positive penalty on the control.

Keywords Optimal experimentation · Value function · Approximation method · Adaptive control · Active learning · Time-varying parameters · Numerical experiments

JEL Classification C63 · E61 · E62

✉ Marco P. Tucci
tucci@unisi.it

Hans M. Amman
amman@uva.nl

[1] Faculty of Economics and Business, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands

[2] Dipartimento di Economia Politica e Statistica, University of Siena, Piazza S. Francesco 7, 53100 Siena, Italy

1 Introduction

In recent years there has been a resurgent interest in economics on the subject of optimal or strategic experimentation also referred to as active learning, see e.g. Amman et al. (2018), Buera et al. (2011) and Savin and Blueschke (2016).[1] There are two prevailing methods for solving this class of models. The first method is based on the value function approach and the second on an approximation method. The former uses dynamic programming for the full problem as used in studies by Prescott (1972), Taylor (1974), Easley and Kiefer (1988), Kiefer (1989), Kiefer and Nyarko (1989), Aghion et al. (1991) and more recently used in the work of Beck and Wieland (2002), Coenen et al. (2005), Levin et al. (2003) and Wieland (2000a, b). A nice set of applications on optimal experimentation, using the value function approach, can be found in Willems (2012).

In principle, the value function approach *should be* the preferred method as it derives the optimal values for the policy variables through Bellman's (1957) dynamic programming. Unfortunately, it suffers from the *curse of dimensionality*, as is shown in Bertsekas (1976). Hence, the value function approach is only applicable to small problems with one or two policy variables. This is caused by the fact that solution space needs to be discretized in such a fashion that it cannot be solved in feasible time. The approximation methods as described in Cosimano (2008) and Cosimano and Gapen (2005a, b), Kendrick (1981) and Hansen and Sargent (2007) use approaches, that are applied in the neighborhood of the linear regulator problems.[2] Because of this *local* nature with respect to the statistics of the model, the method is numerically far more tractable and allows for models of larger dimension. However, the verdict is still out as to how well it performs in terms of approximating the optimal solution derived through the value function. By the way, the approximation method described here, should not be mistaken for a *cautious* or *passive learning* method. Here we concentrate only on optimal experimentation—active learning—approaches.

Both solution methods consider dynamic stochastic models in which the control variables can be used not only to guide the system in desired directions but also to improve the accuracy of estimates of parameters in the models. There is a trade off in which experimentation of the policy variables early in time deviates from reaching current goals, but leads to learning or improved parameter estimates and improved performance of the system later in time. Ergo, the dual nature of the control. For this reason, we concentrate in the sections below on the policy function in the initial period. Usually most of the experimentation—active learning—is done in the beginning of the time interval, and therefore, the largest difference between results obtained with the two methods may be expected in this period.

---

[1] The seminal work on this subject in economics, stems from an early paper by MacRae (1972, 1975), followed by a range of theoretical papers like Easley and Kiefer (1988), Bolton and Harris (1999), Salmon (2001), Moscarini and Smith (2001) and applications like.

[2] For consistency and clarity in the main text, we used the term approximation method instead of adaptive or dual control. The adaptive or dual control approach in MacRae (1975), see Kendrick (1981), Amman (1996) and Tucci (2004), uses methods that draw on earlier work in the engineering literature by Bar-Shalom and Sivan (1969) and Tse (1973).There are differences between this approach and the approximation approaches in Cosimano (2008) and Savin and Blueschke (2016) which we will not discuss in detail here. Through out the paper we will use the approach in Kendrick (1981).

Until very recently there was an invisible line dividing researchers using one approach from those using the other. It is only in Amman et al. (2018) that the value function approach and the approximation method are used to solve the same problem and their solutions are compared. In that paper the focus is on comparing the policy function results reported in Beck and Wieland (2002), through the value function, to those obtained through approximation methods. Therefore those conclusions apply to a situation where the controller is dealing with a nonstationary process and there is no penalty on the control. The goal of this paper is to see if they hold for the more frequently studied case of a stationary process and a positive penalty on the control. To do so a new value function algorithm has been written, to handle several sets of parameters, and more general formulae for the cost-to-go function of the approximation method are used. The remainder of the paper is organized as follows. The problem is stated in Sect. 2. Then the value function approach and the approximation approach are described (Sects. 3 and 4, respectively). Section 5 contains the experiment results. Finally the main conclusions are summarized (Sect. 6).

2 Problem Statement

The discrete time problem we want to investigate dates back to MacRae (1975) and it closely resembles that used in Beck and Wieland (2002). For this reason it is going to be referred to as MBW model throughout the paper. It is defined as follows:
Finding the control that minimize the cost functional, i.e.

$$J = \min_{u_t} \sum_{t=0}^{\infty} \rho^t L(x_t, u_t) \tag{1}$$

subject to the one period loss function taking the form

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

and the linear system

$$x_t = \gamma x_{t-1} + \beta_t u_t + \alpha + \epsilon_t \tag{3}$$

with the time-varying parameter modeled as

$$\beta_t = \beta_{t-1} + \eta_t \tag{4}$$

where $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$ and $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$. The parameter β_t is estimated using the Kalman filter

$$E_{t-1}(\beta_t) = b_{t-1} \tag{5}$$

$$b_t = b_{t-1} + u_t v_{t-1}^b F_t^{-1}(x_t - \alpha - b_{t-1}u_t - \gamma x_{t-1}) \tag{6}$$

$$VAR_{t-1}(\beta_t) = v_{t-1}^b + \sigma_\eta^2 \tag{7}$$

$$v_t^b = v_{t-1}^b - v_{t-1}^b u_t^2 F_t^{-1} v_{t-1} \tag{8}$$

$$F_t = u_t^2 v_{t-1}^b + \sigma_\epsilon^2 \tag{9}$$

The parameters b_0, v_0^b, σ_η^2 and σ_ϵ^2 are assumed to be known.

3 Solving the Value Function

The above problem can be solved using dynamic programming. The corresponding Bellman equation is

$$V(x_{t-1}, u_t) = \min_{u_t} \left\{ L(x_t, u_t | b_t, v_t^b) \right.$$
$$\left. + \rho \int_{-\infty}^{\infty} V(x_t, u_t | b_t, v_t^b) \times f(x_t, u_t | b_t, v_t^b) \, dx_t \right\} \tag{10}$$

with the restrictions in (2)–(9), dropping b_t and v_t^b for convenience, $f(x_t)$ being the normal distribution and ρ the discount factor

$$f(x_t) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp\left[-\left(\frac{x_t - \mu_x}{\sigma_x \sqrt{2}} \right)^2 \right] \tag{11}$$

with mean $E_{t-1}(x_t) = \mu_x$ and $Var_{t-1}(x_t) = \sigma_x^2$, hence

$$V(x_{t-1}, u_t) = \min_{u_t} \left\{ L(x_t, u_t) \right.$$
$$\left. + \rho \int_{-\infty}^{\infty} V(x_t, u_t) \frac{1}{\sigma_x \sqrt{2\pi}} \exp\left[-\left(\frac{x_t - \mu_x}{\sigma_x \sqrt{2}} \right)^2 dx_t \right] \right\} \tag{12}$$

If we use the transform

$$y_t = \frac{x_t - \mu_x}{\sigma_x \sqrt{2}} \tag{13}$$

hence

$$x_t = \sigma_x \sqrt{2} y_t + \mu_x \tag{14}$$

and

$$dx_t = \sigma_x \sqrt{2} dy_t \tag{15}$$

Furthermore

$$\mu_x = E_{t-1}[x_t] = \alpha + b_{t-1}u_t + \gamma x_{t-1} \tag{16}$$

$$\sigma_x^2 = VAR_{t-1}[x_t] = (v_{t-1}^b + \sigma_\eta^2)u_t^2 + \sigma_\epsilon^2 \tag{17}$$

and insert them in (12) we get

$$V(x_{t-1}, u_t) = \min_{u_t} \left\{ L(x_t, u_t) + \frac{\rho}{\sqrt{\pi}} \int_{-\infty}^{\infty} V(x_t, u_t) \times \exp\left[-y_t^2\right] dy_t \right\} \tag{18}$$

The integral part of the right hand side of (18) can be numerically approximated on the $\{y_1 \ldots y_n\}$ nodes with weights $\{\chi_1 \ldots \chi_n\}$ using a gauss-hermite quadrature

$$V(x_{t-1}, u_t) = \min_{u_t} \left\{ L(x_t, u_t) + \frac{\rho}{\sqrt{\pi}} \sum_{k=1}^{n} \chi_k V(x_k, u_t) \right\} \tag{19}$$

x_k the value of x at the node y_k

$$x_k = \sigma_x \sqrt{2} y_k + \mu_x \tag{20}$$

and the necessary updating equations

$$b_k = b_{t-1} + u_t v_{t-1}^b F_t^{-1}(x_k - \alpha - b_{t-1}u_t - \gamma x_{t-1}) \tag{21}$$

$$v_k^b = v_{t-1}^b - v_{t-1}^b u_t^2 F_t^{-1} v_{t-1} \tag{22}$$

We can expand $L(x_t, u_t)$ in (2) as follows[3]

$$\begin{aligned}
L(x_t, u_t) = w\Big[&\alpha^2 + u_t^2(b_{t-1}^2 + v_{t-1}^b + \sigma_\eta^2) + (\gamma x_{t-1} - x^*)^2 + \sigma_\epsilon^2 \\
&+ 2\alpha b_{t-1}u_t + 2\alpha(\gamma x_{t-1} - x^*) + 2b_{t-1}u_t(\gamma x_{t-1} - x^*)\Big] \\
&+ \lambda(u_t - u^*)^2
\end{aligned} \tag{23}$$

The computational challenge is to solve (19) numerically. If we set up a grid

$$x_g \in \{x_{t-1}^{min} \ldots x_{t-1}^{max}\}$$

---

[3] Note that $E_{t-1}(\beta^2) = b_{t-1}^2 + v_{t-1}^b + \sigma_\eta^2$.

of size m_x,

$$b_g \in \{b_{t-1}^{min} \ldots b_{t-1}^{max}\}$$

of size m_b and

$$v_g \in \{v_{t-1}^{b\,min} \ldots v_{t-1}^{b\,max}\}$$

of size m_v,
we can compute an initial guess for V^0 by computing u_t that minimizes $L(x_t, u_t)$ in
Eq. (23) on each of the $m_x \times m_b \times m_v$ tuples $\{b_g, v_g, x_g\}$

$$V_g^0 = \text{argmin}\left[L(x_t, u_t^{CE})\right] \tag{24}$$

where the optimal value of the policy variable u_t^{CE} is equal to

$$u_t^{opt} = -\frac{w(\gamma x_{t-1} - x^*)b_{t-1}}{w(b_{t-1}^2 + v_{t-1}^b + \sigma_\eta^2) + \lambda} - \frac{w\alpha b_{t-1}}{w(b_{t-1}^2 + v_{t-1}^b + \sigma_\eta^2) + \lambda}$$
$$+ \frac{\lambda u^*}{w(b_{t-1}^2 + v_{t-1}^b + \sigma_\eta^2) + \lambda} \tag{25}$$

which is the certainty equivalence (CE) solution of the problem in equations
defined in (1)–(2). Now we have an initial value V^0 we can solve equation (19)
iteratively. [4]

$$V^{j+1} = \underset{u_t}{\text{argmin}}\left\{L(x_t, u_t) + \frac{\rho}{\sqrt{\pi}}\sum_{k=1}^{n}\chi_k V^j(x_k, u_t)\right\} \tag{26}$$

$$b_k = b_{t-1} + u_t v_{t-1}^b F_t^{-1}(x_k - \alpha - b_{t-1}u_t - \gamma x_{t-1}) \tag{27}$$

$$v_k^b = v_{t-1}^b - v_{t-1}^b u_t^2 F_t^{-1} v_{t-1} \tag{28}$$

The value of u_t that minimizes the right hand side of (26) can be obtained through
a simple line search. The value of $V^j(x_k, u_t|b_k, v_k^b)$ in (26), can be found by finding
the corresponding spot on the grid.

---

[4] As far as we know there are no Riccati based recursion methods to solve Eq. (26).

Algorithm
Solving the Bellman equation

Initialization;

    read parameters α, β, ρ, γ, w, λ, x^*, u^*

    setup grid $\{x_{t-1}^{min} \ldots x_{t-1}^{max}\}, \{b_{t-1}^{min} \ldots b_{t-1}^{max}\}, \{v^{b\,min}_{t-1} \ldots v^{b\,max}_{t-1}\}$

    compute $u_t^{CE} = u_t^0$ for each tuple

    compute $V^0(u_t^0)$

while $||V^j - V^{j-1}|| > tol_v$ **do**

    compute $V^j(u_t^j)$

    while $||u_t^j - u_t^{j-1}|| > tol_u$ **do**

        line search to find u_t^{opt}, $\arg\min V(u_t^{opt})$

    end

    $u^{opt} \rightarrow u^{j+1}$;

    $V^{j+1} \rightarrow V^j$

end

4 Approximating the Value Function

In this section we present a short summary of the derivations found in Amman and Tucci (2017). Following Tse and Bar-Shalom (1973) and Kendrick (1981, 2002), for each time period, the approximate cost-to-go at different values of the control is computed. The control yielding the minimum approximate cost is selected. This approximate cost-to-go is decomposed into a deterministic, cautionary and probing component. The deterministic component includes only terms which are not stochastic. The cautionary one includes uncertainty only in the next time period and the probing term contains uncertainty in all future time periods. Thus the probing term includes the motivation to perturb the controls in the present time period in order to reduce future uncertainty about parameter values.

The approximate cost-to-go in the infinite horizon BMW model looks like

$$J_\infty = (\psi_1 + \delta_1)\, u_0^2 + (\psi_2 + \delta_2)\, u_0 + (\psi_3 + \delta_3)$$
$$+ \left(\frac{\sigma_b^2 q}{2}\right) \frac{\phi_1\,(\phi_2 u_0 + \phi_3)^2}{(\sigma_b^2 u_0^2 + q)} \tag{29}$$

Equation (29) is identical to equation (27) in Tucci et al. (2010), but now the parameters associated with the deterministic component, the ψ's, are defined as

$$\psi_1 = \frac{1}{2}\left(\lambda + b_0^2 \rho k^{xx}\right)$$
$$\psi_2 = \rho k^{xx} b_0 \alpha x_0$$
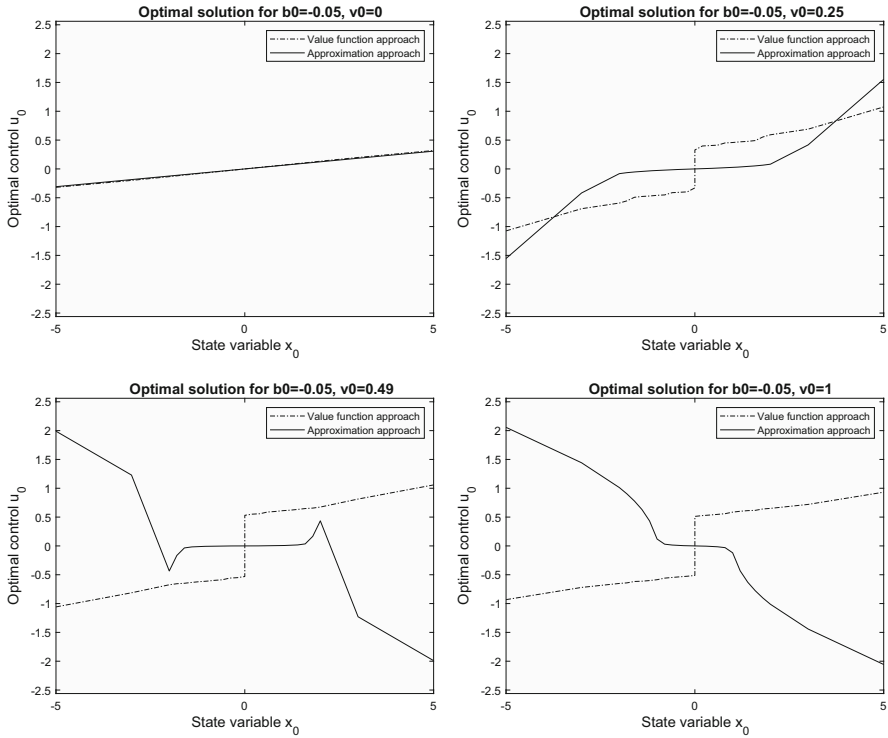$$\psi_3 = \frac{1}{2}\rho k^{xx}\,(\alpha x_0)^2 \tag{30}$$

Fig. 1 Plot for $b_0 = -0.05$

where b_0 is the estimate of the unknown parameter at time 0 and

$$\rho k^{xx} = k_1^{xx} \equiv k^{CE}$$

with k^{CE} the fixed point solution to the usual Riccati recursion[5]

$$k^{CE} = w + \alpha^2 \rho k^{CE} - \left(\alpha \rho k^{CE} b_0\right)^2 \left(\lambda + \rho k^{CE} b_0^2\right)^{-1} \tag{31}$$

The parameters associated with the cautionary component, the δ_i, take the form

$$\delta_1 = \frac{1}{2} v_0^b \left[k_1^{xx} + \tilde{k}_1^{\beta\beta} b_0^2 + 2\tilde{k}_1^{\beta x} b_0 \right]$$

$$\delta_2 = v_0^b \left[\tilde{k}_1^{\beta\beta} b_0 + \tilde{k}_1^{\beta x} \right] \alpha x_0$$

$$\delta_3 = \frac{1}{2} \left[k_1^{xx} q \left(1 - \rho\right)^{-1} + v_0^b \tilde{k}_1^{\beta\beta} \alpha^2 x_0^2 \right] \tag{32}$$

---

[5] In this case the Riccati equation is scalar function and can easily be solved. The multi-dimensional case can be more complicated to solve. See Amman and Neudecker (1997).
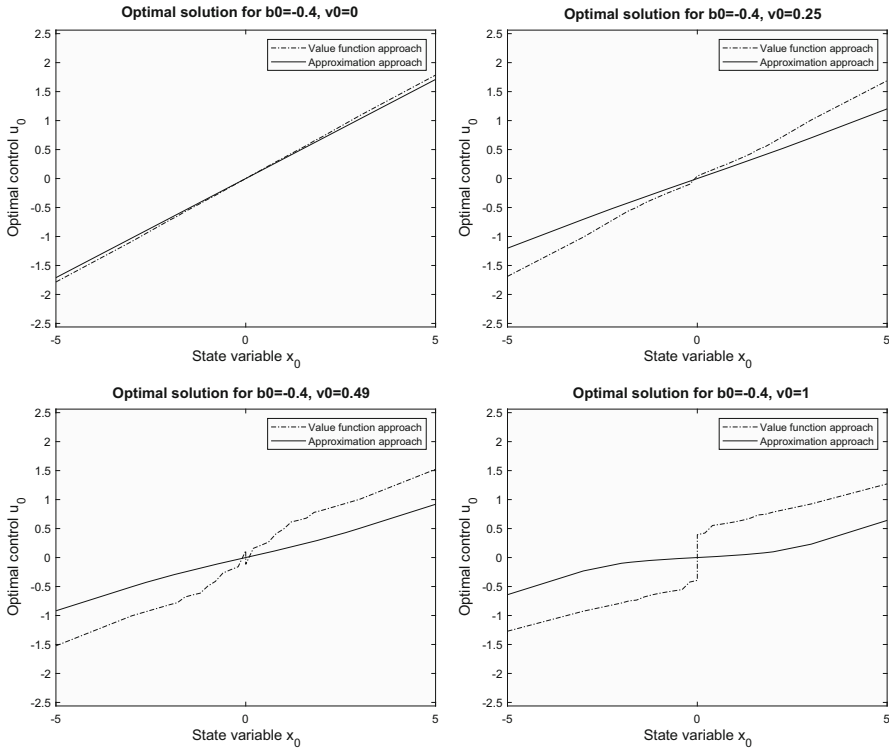
Fig. 2 Plot for $b_0 = -0.4$

with

$$\tilde{k}_1^{\beta x} = 2\rho k_1^{xx} (\alpha + b_0 G) \left[1 - \rho (\alpha + b_0 G)^2 \right]^{-1} G \tag{33}$$

$$\tilde{k}_1^{\beta\beta} = \rho k_1^{xx} \left[1 + 3\rho (\alpha + b_0 G)^2 \right] \left[1 - \rho (\alpha + b_0 G)^2 \right]^{-2} G^2$$

$$- \left(\rho k_1^{xx} \right)^2 \left\{ \alpha + 2b_0 G \left[1 - \rho (\alpha + b_0 G)^2 \right]^{-1} \right\}^2$$

$$\left(\lambda_1 + \rho k_1^{xx} b_0^2 \right)^{-1} \left[1 - \rho (\alpha + b_0 G)^2 \right]^{-1} \tag{34}$$
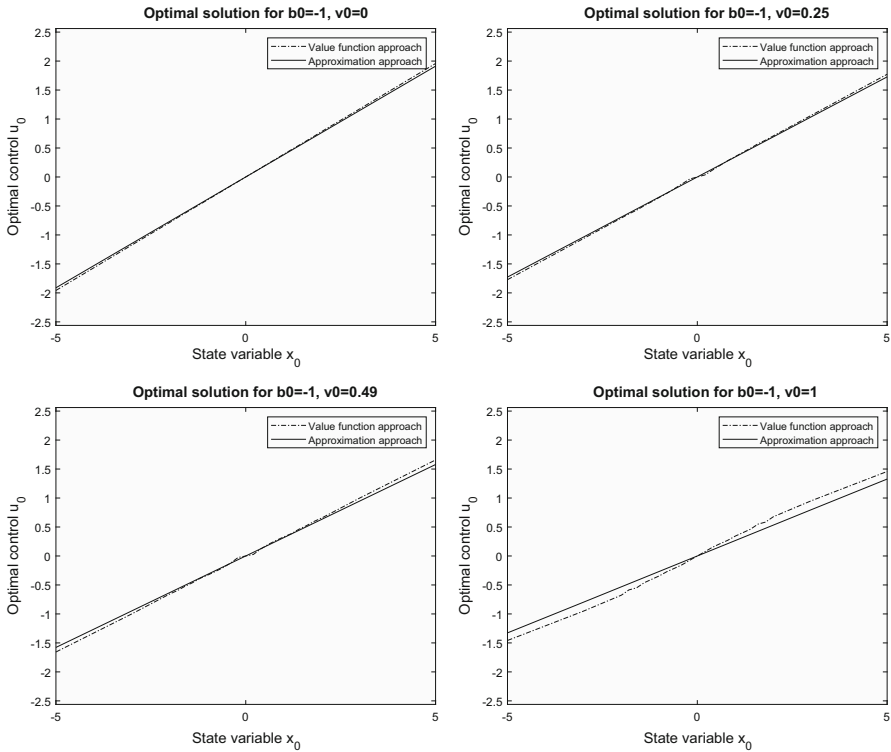
Fig. 3 Plot for $b_0 = -1.0$

where[6]

$$G = - \left(\lambda + \rho k^{CE} b_0^2 \right)^{-1} \alpha \rho k^{CE} b_0 \qquad (35)$$

Finally the parameters related to the probing component, the ϕ's, take the form

$$\phi_1 = \left[\left(\rho k_1^{xx} \right)^2 \left(\lambda_1 + \rho k_1^{xx} b_0^2 \right)^{-1} \right] \left[1 - \rho \left(\alpha + b_0 G \right)^2 \right]^{-2}$$

$$\phi_2 = \left\{ \alpha + 2 b_0 G \left[1 - \rho \left(\alpha + b_0 G \right)^2 \right]^{-1} \right\} b_0$$

---

[6] This compares with $\tilde{k}_1^{\beta x} = 2 w_2 \left(\alpha + b G_1 \right) G_1$ and

$$\tilde{k}_1^{\beta \beta} = w_2 G_1^2 + w_2^2 \left(\alpha + 2 b G_1 \right)^2 \left[- \left(\lambda_1 + b^2 w_2 \right) \right]^{-1}$$

where the feedback matrix is defined as $G_1 = \left(-\alpha b w_2 / \lambda_1 + b^2 w_2 \right)$, in the two-period finite horizon model.
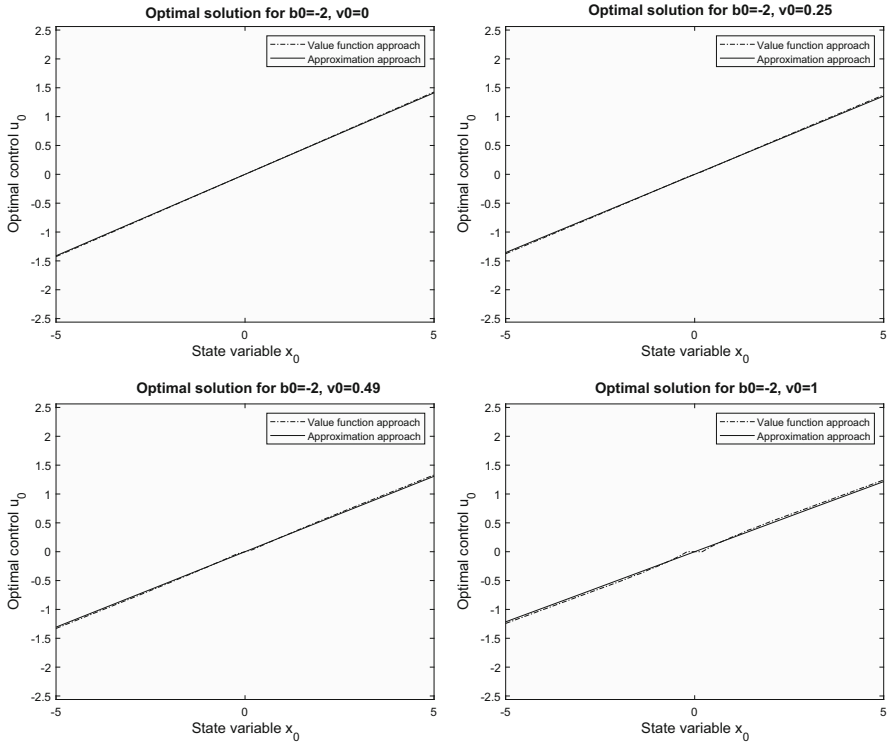
Fig. 4 Plot for $b_0 = -2.0$

$$\phi_3 = \left\{ \alpha + 2b_0 G \left[1 - \rho \left(\alpha + b_0 G \right)^2 \right]^{-1} \right\} \alpha x_0 \qquad (36)$$

As shown in Amman and Tucci (2017) the new definitions are perfectly consistent with those associated to the two-period finite horizon model.

5 Experimentation

In this section the infinite horizon control for the MBW model is computed for the value function and approximation method when the system is assumed stationary. Moreover an equal penalty weight is applied to deviations of the state an control from their desired path, assumed zero here. In order to stay as close as possible to the case discussed in Beck and Wieland (2002, p. 1367) and Amman et al. (2018) the parameters are $\alpha = 0.7$, $\gamma = 0$, $q = 1$, $w = 1$, $\lambda = 1$, $\rho = 0.95$.

Figures 1, 2, 3 and 4 contain the four typical solutions of the model for $b_0 = -0.05$, $b_0 = -0.4$, $b_0 = -1.0$ and $b_0 = -2.0$. For each value of b_0, four different levels of uncertainty, as measured by the estimated parameter variance v_0, are considered. By doing so it is possible to assess the relevance of relative uncertainty on the performance of the two approaches considered in this work. In this situation both the
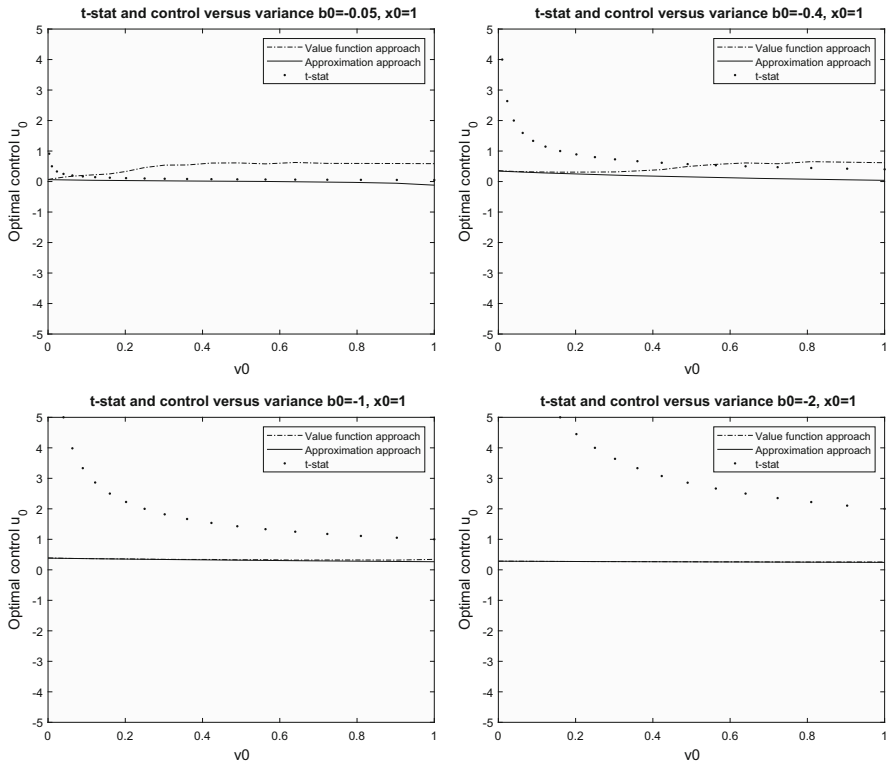
Fig. 5 Comparison control versus variance and *t*-statistics

approximation approach (solid line) and the value function approach (dotted line) suggest a more conservative control than in the nonstationary and no penalty on the control case. The difference between the two approaches tends to be much smaller when the initial state is not too far from the desired path and the parameter uncertainty is relatively small, whereas it is approximately the same for $x_0 = -5$ or $x_0 = 5$ (compare Figure (1) in Amman et al. (2018) with the top right panel in Fig. 2). [7] By comparing the different cases reported below, it is apparent that the difference between the solutions generated by the two methods depends heavily upon the level of relative uncertainty about the unknown parameter. For instance in Fig. 2, top-left panel , it is shown that the two approaches generate almost identical results for the whole spectrum of possible initial states when b_0 is estimated with very little variance, i.e. 0.01. The situation changes, Fig. 2 remaining panels, when the same parameter estimate is associated with higher variances. A small shock from the desired initial state causes the two algorithms to give meaningfully different solutions when the variance is 1. Therefore it turns out that in all cases, Figs. 1, 2 and 3, relative uncertainty is very important in explaining the different results. Moreover it turns out that the distinction between high relative uncertainty

---

[7] The reader should keep in mind that the opposite convention is used in Amman et al. (2018).
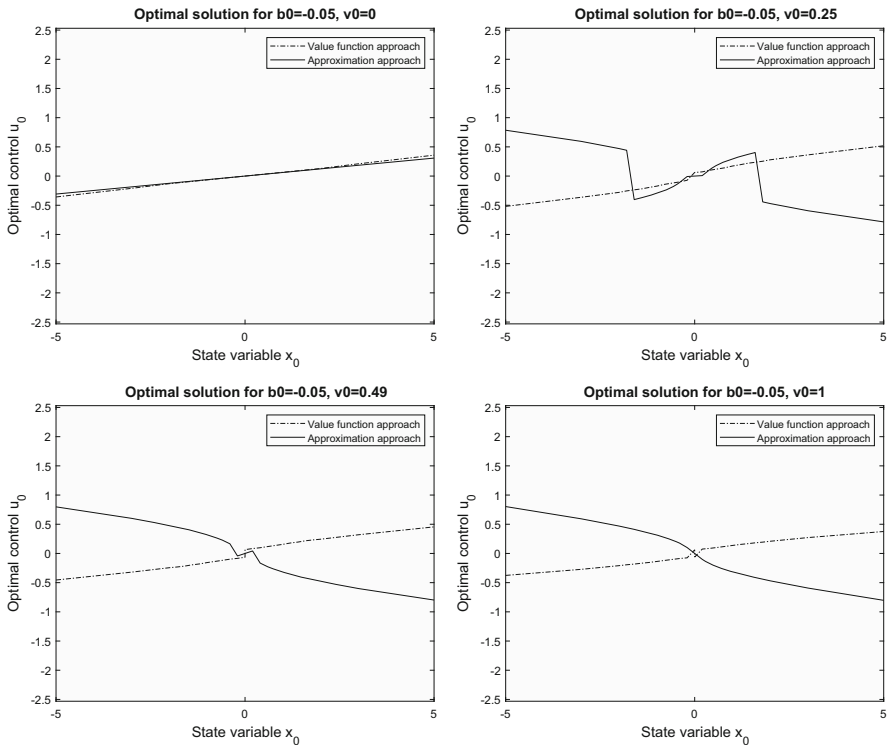
Fig. 6 Plot for $b_0 = -0.05$

and extreme uncertainty becomes relevant. For instance in Fig. 1, bottom panels, the unknown parameter is extremely uncertain and the two approaches suggest *opposite* controls.

A readily availble measure of relative uncertainty of an unknown parameter is the usual *t-statistics* automatically generated by all econometrics software, i.e. the ratio of the parameter estimate over its estimated standard deviation. By using this econometric tool it can be stated that when there is very little or no uncertainty about the unknown parameter as in Fig. 4, a situation where the t -statistics ranges from virtual certainty (top left panel) to 2 (bottom right panel), the two solutions are almost identical as it should be expected. As the level of uncertainty increases, as in Figs. 2 and 3, the difference becomes more pronounced and the approximation method is usually less active then the value function approach. Figure 2, with the t ranging from certainty to 0.4, and 3, with t going from certainty to 1, reflect the most common situations. However when there is high uncertainty as in Fig. 1, where the t goes from 5 (top left panel) to 0.05 (bottom right panel), the approximation method shows very aggressive solutions when the t -statistics is around 0.1–0.2 and the initial state is far from its desired path. In the extreme cases where the t drops below 0.1, bottom panels of Fig. 1, this method finds optimal to perturb the system in the 'opposite' direction in order to learn something about the the unknown parameter. These are cases where the
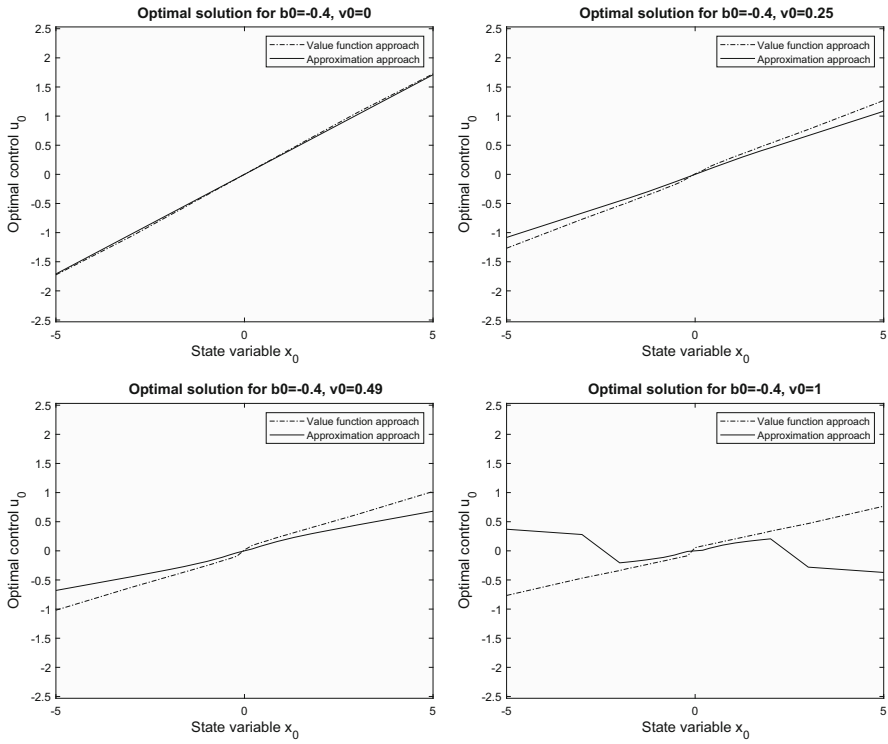
Fig. 7 Plot for $b_0 = -0.4$

99 percent confidence intervals for the unknown parameter are $(-2.15:2.05)$, when $v_0 = 0.49$, and $(-3.05:2.95)$, when $v_0 = 1$. Alternatively, if the initial state is close to the desired path this method is very conservative.

On the other hand the value function approach seems somehow 'insulated' by the extreme uncertainty surrounding the unknown parameter. As apparent from Fig. 1 this optimal control stays more or less constant in the presence of an extremely uncertain parameter. The major consequence seems to be a bigger 'jump' in the control applied when the initial state is around the desired path. Summarizing, a very higher parameter uncertainty results in a more aggressive control when the initial state is in the neighborhood of its desired path and a relatively less aggressive control when it is far from it.

Figures 1, 2, 3 and 4 show the importance of the parameter relative uncertainty, as measured by its *t-statistics*, in explaining the difference between the two alternative approaches. But is there a continuous monotonous relationship? At what levels of this t-value the differences become pronounced? Is there a threshold value under which 'extreme uncertainty' results appear? To answer these questions.

Figure 5 uses the same four values of b_0 to compare the two methods at various variances, when the initial state is $x_0 = 1$ explicitly taking into account the *t-ratio*. It is apparent that there is an inverse relationship between this statistics and the discrep-
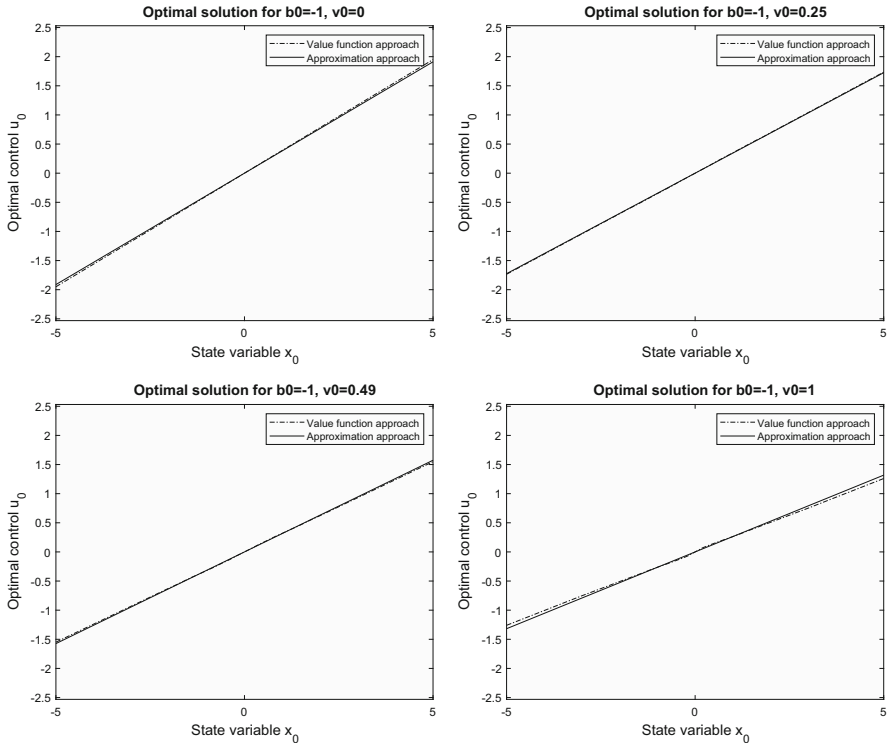
Fig. 8 Plot for $b_0 = -1.0$

ancy between the two approaches under consideration. Again the difference is more noticeable when the *t-statistics* drops below 1. In the presence of extreme uncertainty, i.e. when this statistics falls below 0.5, and an initial state far from the desired path this difference not only gets larger and larger but it may be also associated with the approaches giving *opposite* solutions, i.e. a positive control vs a negative control. This is what happens in the top left panel of Fig. 5 where for very low *t-statistics* the value function approach suggests a positive control whereas the approximation approach suggests a slightly negative control.

The same qualitative results characterize a situation with a much smaller system variance, namely $q = 0.01$, as shown in Figs. 6, 7, 8 and 9. In this scenario controls are less aggressive than in the previous one and, as previously seen, the approximation approach is generally less active than the competitor. It looks like the optimal control is insensitive to system noise when the parameter associated with it has very little uncertainty as in the top left panel of Figs. 1 and 6. However, when the unknown parameter has a very low *t*-statistics the control is significantly affected by the system noise. Then the distinction between high and extreme uncertainty about the unknown parameter becomes even more relevant then before. At a preliminary examination it seems that a higher system noise has the effect of 'reducing' the perceived parameter uncertainty. For example, the bottom right panels in Figs. 2 and 7 show the opti-
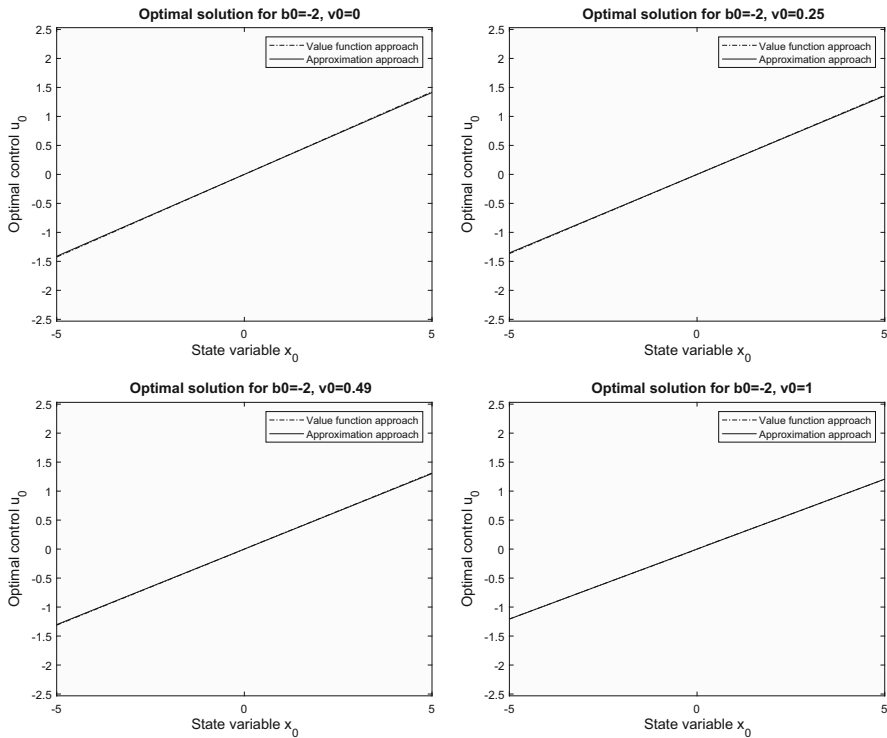
Fig. 9 Plot for $b_0 = -2.0$

mal controls when the t associated with the unknown parameter is around 0.5. This parameter uncertainty is associated with a very low system noise in the latter case. Therefore it is perceived in its real dimension and the approximation approach suggests a control in the 'opposite' direction when the initial state is far from its desired path.

Figure 10 uses the same four values of b_0 to compare the two methods at various variances, when the initial state is $x_0 = 1$ and the system variance is $q = 0.1$. As in Fig. 10 the difference is more noticeable when the t-statistics drops below 1.

It is unclear at this stage if the distinction between high uncertainty and extreme uncertainty is relevant also for the nonstationary case treated in Amman et al. (2018). A hint may be given by their Fig. 8. It reports the results for the case where the parameter estimate is 0.3 and its variance is 0.49, i.e. the *t-statistics* of the unknown parameter is around 0.4. In this case the approximation approach is more active than the value function approach when the initial state is far from the desired path, i.e. x_0 greater than 3. This seems to suggest that the distinction between high and extreme uncertainty is relevant also when the system is nonstationary and no penalty is applied to the controls.
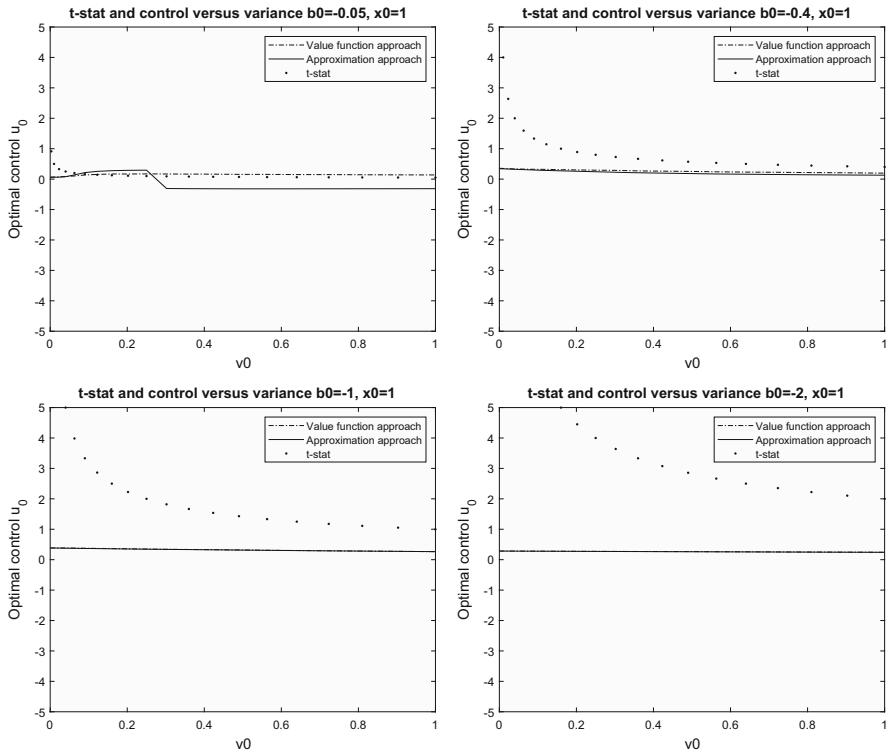
Fig. 10 Comparison control versus variance and *t*-statistics

6 Conclusions

In a previous paper Amman et al. (2018) compare the value function and the approximation method in a situation where the controller is dealing with a nonstationary process and there is no penalty on the control. They conclude that differences are small when the effects of learning are limited. In this paper we find that similar results hold for the more commonly studied case of a controller facing a stationary process and a positive penalty on the control. Moreover we find that a good proxy for parameter uncertainty is the usual *t* -statistics used by econometricians and that it is very important to distinguish between high and in extreme uncertainty about the unknown parameter. In the latter situation, i.e. *t* close to 0, when the initial state is very far from its desired path and the parameter associated with the control is very small the approximation method becomes very active. Eventually it even perturbs the system in the *opposite* direction.

This is something that needs further investigation with other models and parameter sets. It may be due to the fact that the computational approximation to the integral needed in value function approach does not fully incorporate these extreme cases. Or it may the consequence of some hidden relationships between the parameters and the components of the cost-to-go in the approximation approach. However the behavior

of the 'approximation control' makes full sense. Its suggestion is 'in the presence of extreme uncertainty don't be very active if you are close to the desired path but 'go wild' if you are far from it'. If this characteristics is confirmed it may represent a useful additional tool in the hands of the control researcher to discriminate between cases where the control can be reliably applied and cases where it cannot.

References

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

Amman, H., & Tucci, M. (2017). The dual approach in an infinite horizon model. Quaderni del Dipartimento di Economia Politica 766. Università di Siena, Siena, Italy.

Amman, H. M. (1996). Numerical methods for linear-quadratic models. In H. M. Amman, D. A. Kendrick, & J. Rust (Eds.), *Handbook of computational economics of handbook in economics* (Vol. 13, pp. 579–618). Amsterdam: North-Holland Publishers.

Amman, H. M., Kendrick, D. A., & Tucci, M. P. (2018). Approximating the value function for optimal experimentation. *Macroeconomic Dynamics* (forthcoming).

Amman, H. M., & Neudecker, H. (1997). Numerical solution methods of the algebraic matrix riccati equation. *Journal of Economic Dynamics and Control*, *21*, 363–370.

Bar-Shalom, Y., & Sivan, R. (1969). On the optimal control of discrete-time linear systems with random parameters. *IEEE Transactions on Automatic Control*, *14*, 3–8.

Beck, G., & Wieland, V. (2002). Learning and control in a changing economic environment. *Journal of Economic Dynamics and Control*, *26*, 1359–1377.

Bellman, R. E. (1957). *Dynamic programming*. Princeton, NJ: Princeton University Press.

Bertsekas, D. P. (1976). *Dynamic programming and stochastic control* (Vol. 125)., Mathematics in science and engineering New York: Academic Press.

Bolton, P., & Harris, C. (1999). Strategic experimentation. *Econometrica*, *67*(2), 349–374.

Buera, F. J., Monge-Naranjo, A., & Primiceri, G. E. (2011). Learning the wealth of nations. *Econometrica*, *79*(1), 1–45.

Coenen, G., Levin, A., & Wieland, V. (2005). Data uncertainty and the role of money as an information variable for monetary policy. *European Economic Review*, *49*, 975–1006.

Cosimano, T. F., & Gapen, M. T. (2005a). Program notes for optimal experimentation and the perturbation method in the neighborhood of the augmented linear regulator problem. Working paper. Notre Dame, IN: Department of Finance, University of Notre Dame.

Cosimano, T. F., & Gapen, M. T. (2005b). Recursive methods of dynamic linear economics and optimal experimentation using the perturbation method. Working paper. Notre Dame, IN: Department of Finance, University of Notre Dame.

Cosimano, T. F. (2008). Optimal experimentation and the perturbation method in the neighborhood of the augmented linear regulator problem. *Journal of Economics, Dynamics and Control*, *32*, 1857–1894.

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

Hansen, L. P., & Sargent, T. J. (2007). *Robustness*. Princeton, NJ: Princeton University Press.

Kendrick, D. A. (2002). Stochastic control for economic models, online edn. http://www.utexas.edu/cola/depts/economics/faculty/dak2.

Kendrick, D. A. (1981). *Stochastic control for economic models* (2nd ed., p. 2002). New York, NY: McGraw-Hill Book Company.

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.

Levin, A., Wieland, V., & Williams, J. C. (2003). The performance of forecast-based monetary policy rules under model uncertainty. *American Economic Review*, *93*, 622–645.

MacRae, E. C. (1972). Linear decision with experimentation. *Annals of Economic and Social Measurement*, *1*, 437–448.

MacRae, E. C. (1975). An adaptive learning role for multi-period decision problems. *Econometrica*, *43*, 893–906.

Moscarini, G., & Smith, L. (2001). The optimal level of experimentation. *Econometrica*, *69*(6), 1629–1644.

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

Salmon, T. C. (2001). An evaluation of econometric models of adaptive learning. *Econometrica*, *69*(6), 1597–1628.

Savin, I., & Blueschke, D. (2016). Lost in translation: Explicitly solving nonlinear stochastic optimal control problems using the median objective value. *Computational Economics*, *48*, 317–338.

Taylor, J. B. (1974). Asymptotic properties of multi-period control rules in the linear regression model. *International Economic Review*, *15*, 472–482.

Tse, E. (1973). Further comments on adaptive stochastic control for a class of linear systems. *IEEE Transactions on Automatic Control*, *18*, 324–326.

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

Tucci, M. P. (2004). *The rational expectation hypothesis, time-varying parameters and adaptive control*. Dordrecht: Springer.

Tucci, M. P., Kendrick, D. A., & Amman, H. M. (2010). The parameter set in an adaptive control Monte Carlo experiment: Some considerations. *Journal of Economic Dynamics and Control*, *34*, 1531–1549.

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.

Willems, T. (2012). *Essays on Optimal Experimentation*. PhD thesis, Tinbergen Institute, University of Amsterdam, Amsterdam.