



UNIVERSITÀ
DI SIENA
1240

**QUADERNI DEL DIPARTIMENTO
DI ECONOMIA POLITICA E STATISTICA**

**Hans M Amman
Marco P. Tucci**

How active is active learning: value function
method vs an approximation method

n. 788 – Ottobre 2018



How active is active learning: value function method vs an approximation method

Hans M. Amman^{a,*}, Marco P. Tucci^b

^a*Faculty of Economics and Business, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, the Netherlands*

^b*Marco P. Tucci, Dipartimento di Economia Politica e Statistica, Università di Siena, Piazza San Francesco 7, 53100 Siena, Italy*

Abstract

In a previous paper Amman and Tucci (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 (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.

*Corresponding author.

Date: 19/10/2018, Time: 14:38, Document: howactive.tex

Email addresses: amman@uva.nl (Hans M. Amman), marco.tucci@unisi.it (Marco P. Tucci)

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 and Tucci (2018), Buera *et al.* (2011), 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; 2000b). 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 de policy variables through Bellman's (1957) dynamic programming. Unfortunately, it suffers from the *curse of dimensionality*, Bertsekas (1976), and 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; 2005b), 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,

¹The 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.

²For 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).

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. Thus, there is a trade off in which experimentation of the policy variables early in time detracts from reaching current goals, but leads to learning or improved parameter estimates and thus 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.

Until very recently there was an invisible line dividing researchers using one approach from those using the other. It is only in Amman and Tucci (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 (Amman and Tucci (2018)). The remainder of the paper is organized as follows. The problem is stated in Section 2. Then the value function approach and the approximation approach are described (Section 3 and 4, respectively). Section 5 contains the experiment results. Finally the main conclusions are summarized (Section 6).

2. Problem statement

The 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

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

subject to

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

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

$$\beta_t = \beta_{t-1} + \eta_t \quad (4)$$

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 a Kalman filter

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

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

$$\text{VAR}_{t-1}(\beta_t) = v_{t-1}^b + \sigma_\eta^2 \quad (7)$$

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

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

The parameters $b_0, v_0^b, \sigma_\eta^2$ and σ_ϵ^2 are assumed to be known.

3. Solving the Value Function

The above problem can be solve used dynamic programming. The corresponding Bellman equation is

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

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

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

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

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

If we use the transform

$$y_t = \frac{x_t - \mu_x}{\sigma_x \sqrt{2}} \quad (13)$$

hence

$$x_t = \sigma_x \sqrt{2} y_t + \mu_x \quad (14)$$

and

$$dx_t = \sigma_x \sqrt{2} dy_t \quad (15)$$

Furthermore

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

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

and insert them in (12) we get

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

The integral part of the right hand side of (18) can be numerically approximated on the $\{y_1 \dots y_n\}$ nodes with weights $\{\omega_1 \dots \omega_n\}$ using a gauss-hermite quadrature

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

x_k the value of x at the node y_k

$$x_k = \sigma_x \sqrt{2} y_k + \mu_x \quad (20)$$

and the necessary updating equations

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

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

We can expand $L(x_t, u_t)$ in (2) as follows³

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

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

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

of size m_x ,

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

of size m_b and

$$v_g \in \{v_{t-1}^{bmin} \dots v_{t-1}^{bmax}\}$$

of size m_v ,

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

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

$$V_g^0 = \operatorname{argmin}[L(x_t, u_t^{CE})] \quad (24)$$

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

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

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

$$V^{j+1} = \operatorname{argmin}_{u_t} \left\{ L(x_t, u_t) + \frac{\rho}{\sqrt{\pi}} \sum_{k=1}^n \omega_k V^j(x_k, u_t) \right\} \quad (26)$$

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

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

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.

Algorithm Solving the Bellman equation

Initialization;

```

    read parameters  $\alpha, \beta, \rho, \gamma, w, \lambda, x^*, u^*$ 
    setup grid  $\{x_{t-1}^{min} \dots x_{t-1}^{max}\}, \{b_{t-1}^{min} \dots b_{t-1}^{max}\}, \{v_{t-1}^{bmin} \dots v_{t-1}^{bmax}\}$ 
    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  $u_t^{opt}, \text{argmin} V(u_t^{opt})$ 
    | 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). The approximate cost-to-go in the infinite horizon BMW model looks like

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

Equation (29) is identical to equation (5.5) in Tucci et al. (2010), but now the parameters associated with the deterministic component, the ψ 's, are defined as

$$\begin{aligned} \psi_1 &= \frac{1}{2} (\lambda + b_0^2 \rho k^{xx}) \\ \psi_2 &= \rho k^{xx} b_0 \alpha x_0 \\ \psi_3 &= \frac{1}{2} \rho k^{xx} (\alpha x_0)^2 \end{aligned} \quad (30)$$

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 recursion⁴

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

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

$$\begin{aligned} \delta_1 &= \frac{1}{2} \nu_0^b \left[k_1^{xx} + \tilde{k}_1^{\beta\beta} b_0^2 + 2\tilde{k}_1^{\beta x} b_0 \right] \\ \delta_2 &= \nu_0^b \left[\tilde{k}_1^{\beta\beta} b_0 + \tilde{k}_1^{\beta x} \right] \alpha x_0 \\ \delta_3 &= \frac{1}{2} \left[k_1^{xx} q (1 - \rho)^{-1} + \nu_0^b \tilde{k}_1^{\beta\beta} \alpha^2 x_0^2 \right] \end{aligned} \quad (32)$$

with

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

$$\begin{aligned} \tilde{k}_1^{\beta\beta} &= \rho k_1^{xx} \left[1 + 3\rho (\alpha + b_0 G)^2 \right] \left[1 - \rho (\alpha + b_0 G)^2 \right]^{-2} G^2 \\ &\quad - (\rho k_1^{xx})^2 \left\{ \alpha + 2b_0 G \left[1 - \rho (\alpha + b_0 G)^2 \right]^{-1} \right\}^2 \\ &\quad \left(\lambda_1 + \rho k_1^{xx} b_0^2 \right)^{-1} \left[1 - \rho (\alpha + b_0 G)^2 \right]^{-1} \end{aligned} \quad (34)$$

where⁵

⁴In 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).

⁵This compares with $\tilde{k}_1^{\beta x} = 2w_2 (\alpha + bG_1) G_1$ and

$$\tilde{k}_1^{\beta\beta} = w_2 G_1^2 + w_2^2 (\alpha + 2bG_1)^2 \left[-(\lambda_1 + b^2 w_2) \right]^{-1}$$

where the feedback matrix is defined as $G_1 = (-\alpha b w_2 / \lambda_1 + b^2 w_2)$, in the two-period finite horizon model.

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

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

$$\begin{aligned} \phi_1 &= \left[(\rho k_1^{xx})^2 \left(\lambda_1 + \rho k_1^{xx} b_0^2 \right)^{-1} \right] \left[1 - \rho (\alpha + b_0 G)^2 \right]^{-2} \\ \phi_2 &= \left\{ \alpha + 2b_0 G \left[1 - \rho (\alpha + b_0 G)^2 \right]^{-1} \right\} b_0 \\ \phi_3 &= \left\{ \alpha + 2b_0 G \left[1 - \rho (\alpha + b_0 G)^2 \right]^{-1} \right\} \alpha x_0 \end{aligned} \quad (36)$$

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 and 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, page 1367) and Amman *et al.* (2018) the parameters are $\alpha = 0.7, \gamma = 0, q = 1, w = 1, \lambda = 1, \rho = 0.95$.

Figures (1-4) contain the four typical solutions of the model for $b_0 = -0.05, b_0 = -0.4, b_0 = -1.0$ and $b_0 = -2.0$. In this situation both the 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 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 Figure 2). The reader should keep in mind that the opposite convention is used in Amman and Tucci (2018). 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 uncertainty

about the unknown parameter. Moreover it turns out that the distinction between high uncertainty and extreme uncertainty becomes relevant.

Figure 1:
Plot for $b_0 = -0.05$

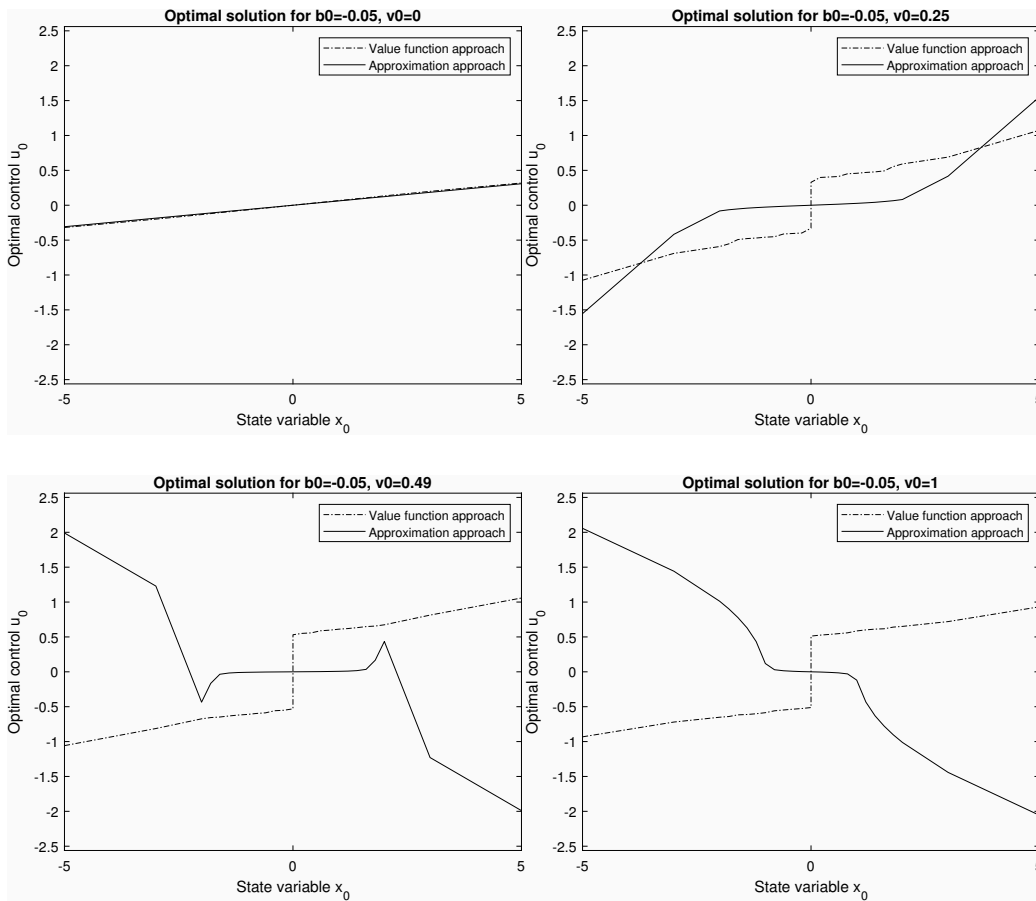
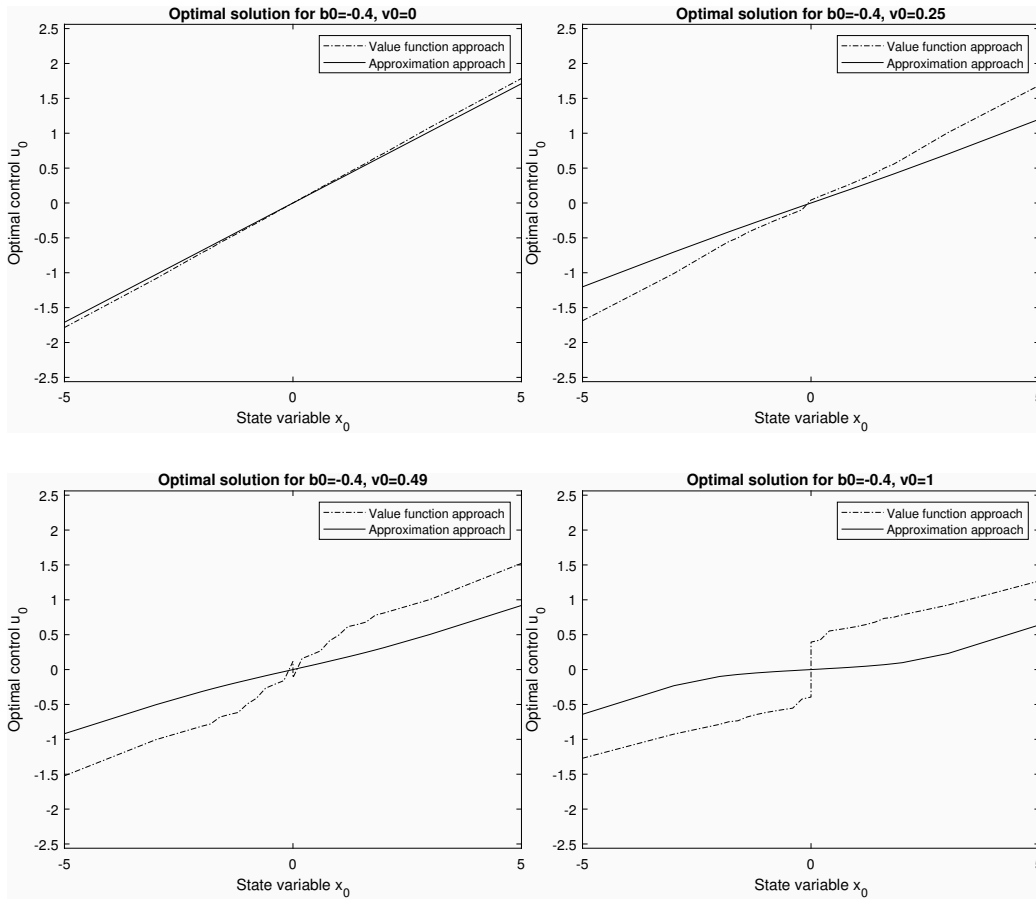


Figure 2:
Plot for $b_0 = -0.4$



When there is very little or no uncertainty about the unknown parameter as in Figure 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 Figure 3 and 2, the difference becomes more pronounced and the approximation method is usually less active than 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 Figure 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 Figure 1, this method finds optimal to perturb the system in the 'opposite' direction in order to learn something about the unknown parameter. These are cases where the 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 Figure 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 high 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.

Figure 3:
Plot for $b_0 = -1.0$

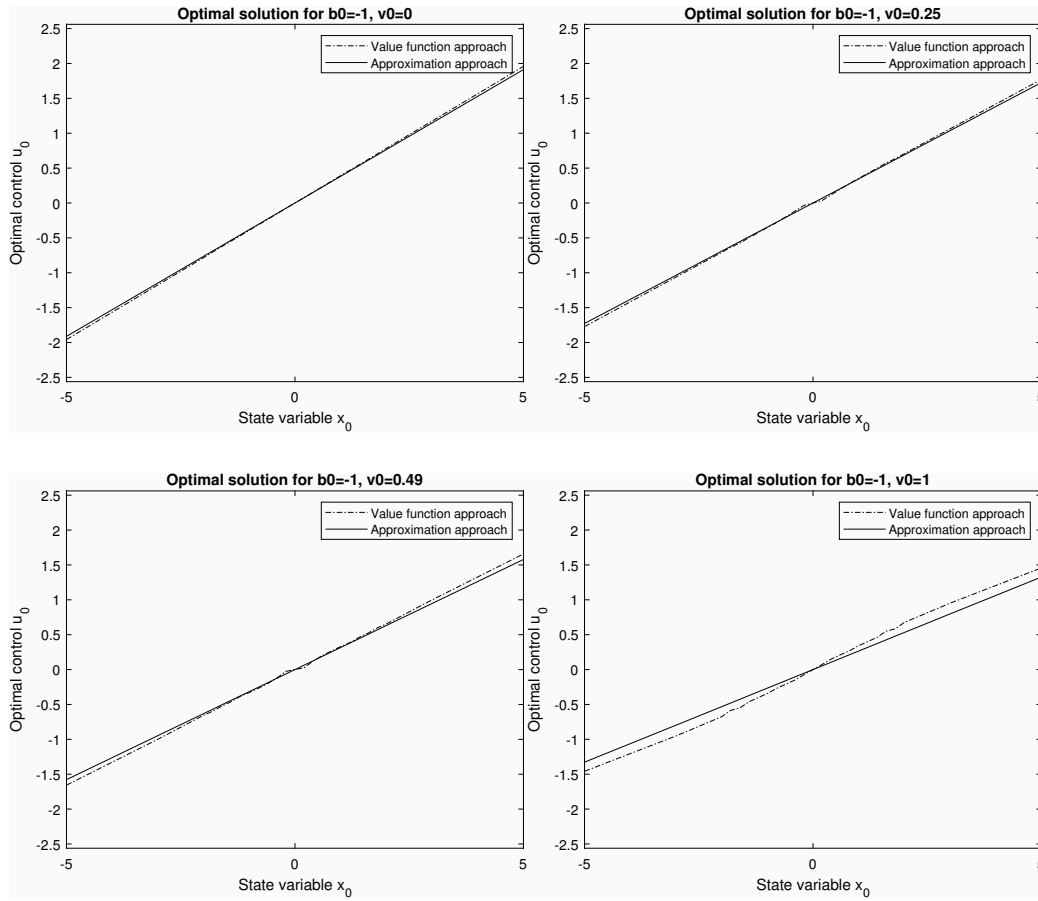


Figure 4:
Plot for $b_0 = -2.0$

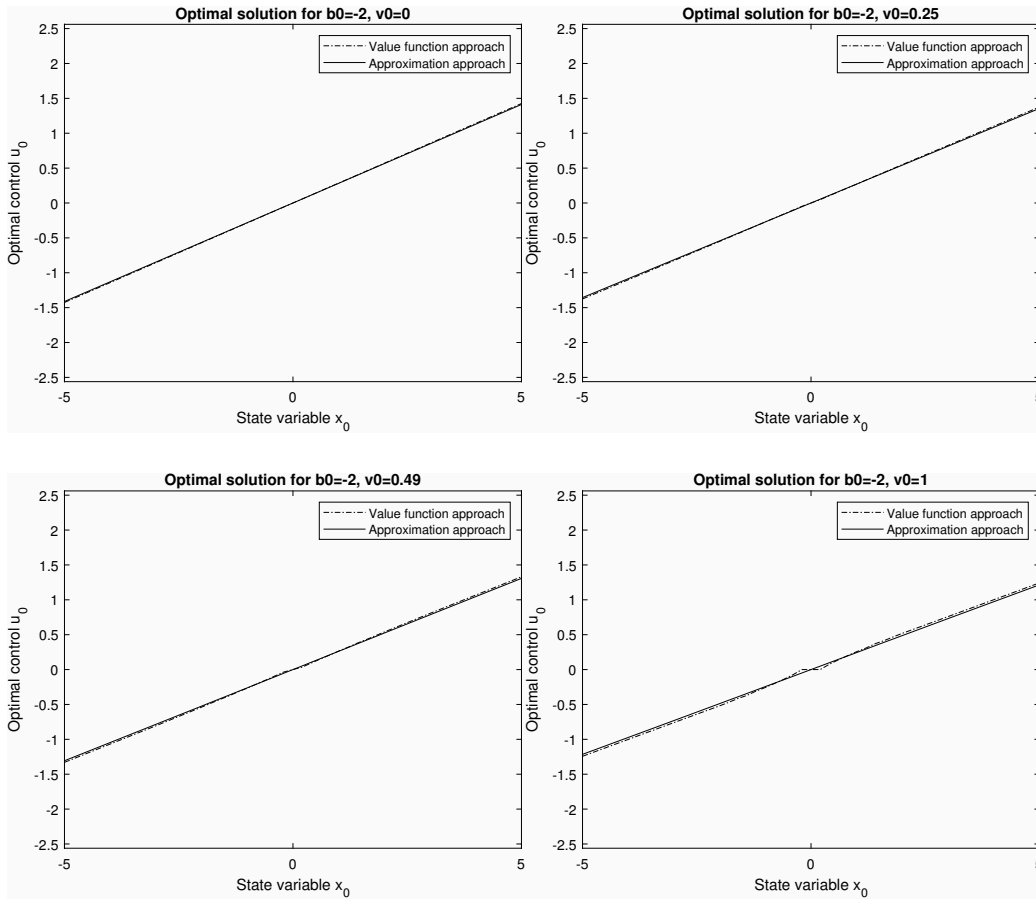
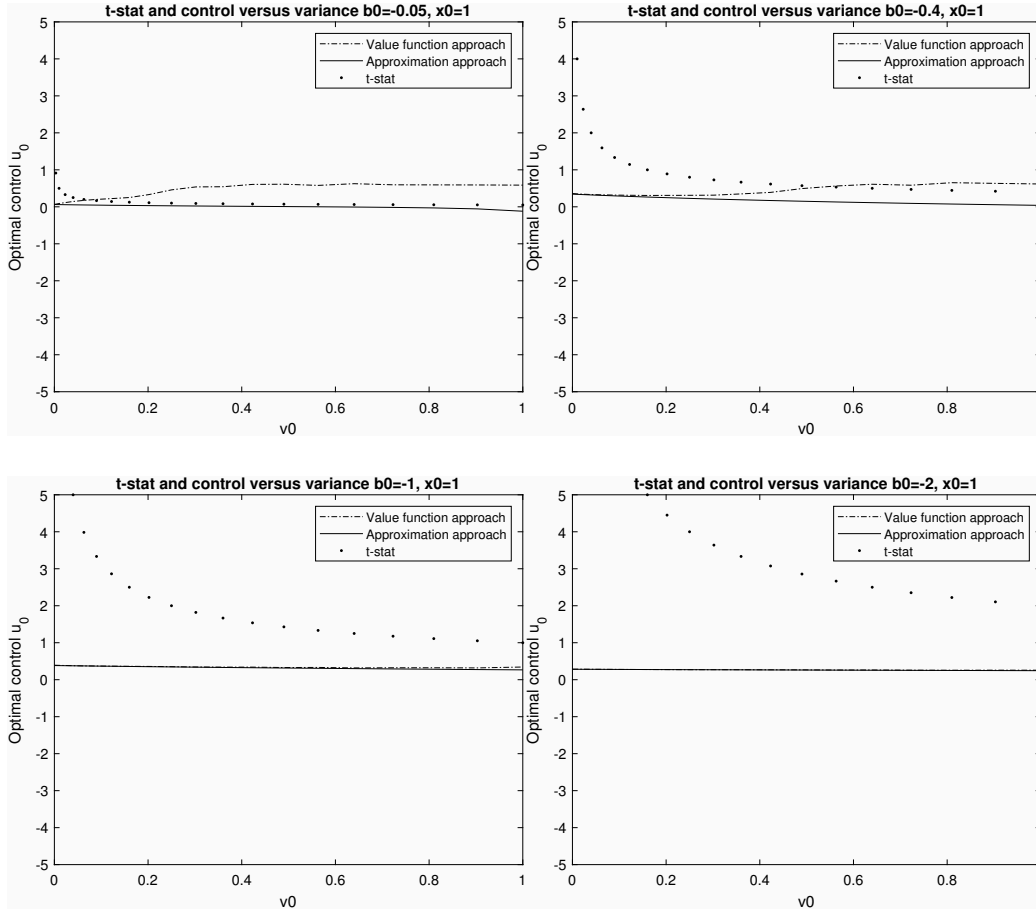


Figure 5 uses the same four values of b_0 to compare the two methods at various variances, when the initial state is $x_0 = 1$. Again the difference is more noticeable when the t -statistics drops below 1. In the presence of extreme uncertainty, i.e. when this statistics falls below 0.5, and an initial 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 Figure 5 where for very low t -statistics the value function approach suggests a positive control whereas the approximation approach suggests a slightly negative control.

Figure 5: Comparison Control versus Variance and t -statistics



The same qualitative results characterize a situation with a much smaller system variance, namely $q = 0.01$, as shown in Figures (6-9). 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 Figures 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 than 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 Figure 2 and 7 show the optimal 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 6:
Plot for $b_0 = -0.05$

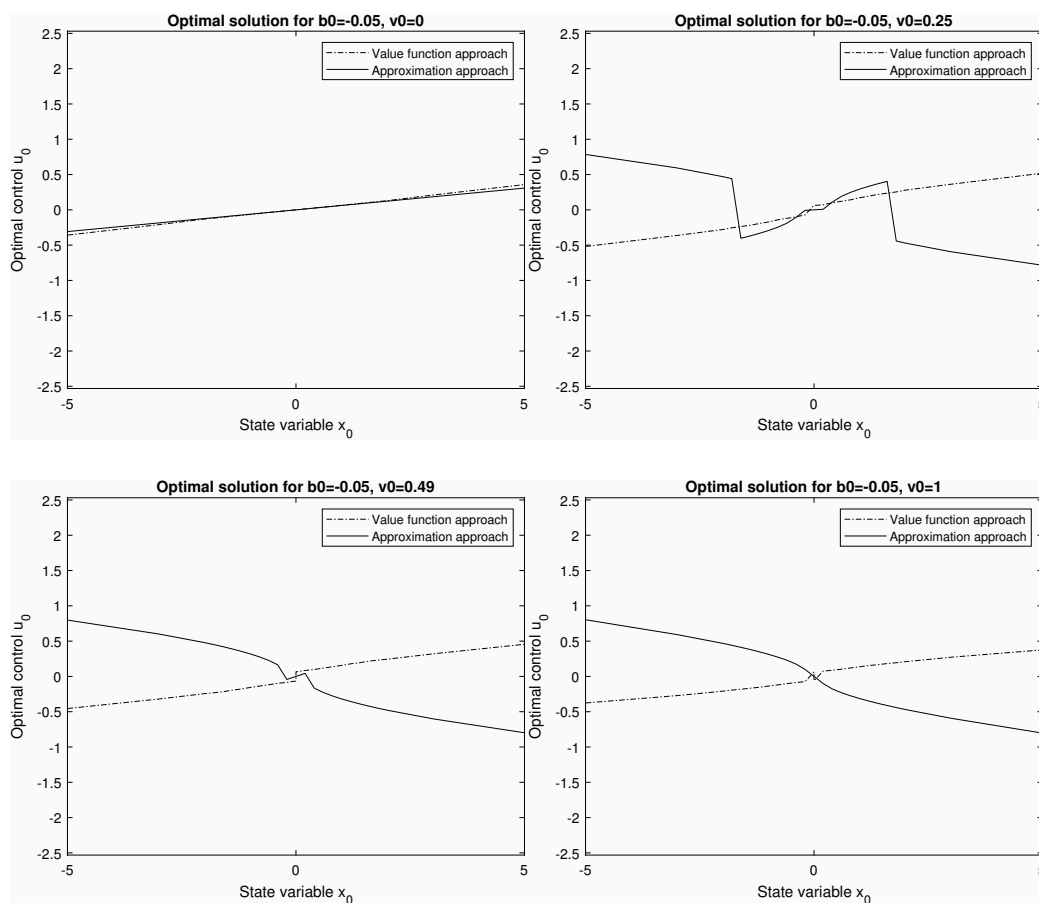


Figure 7:
Plot for $b_0 = -0.4$

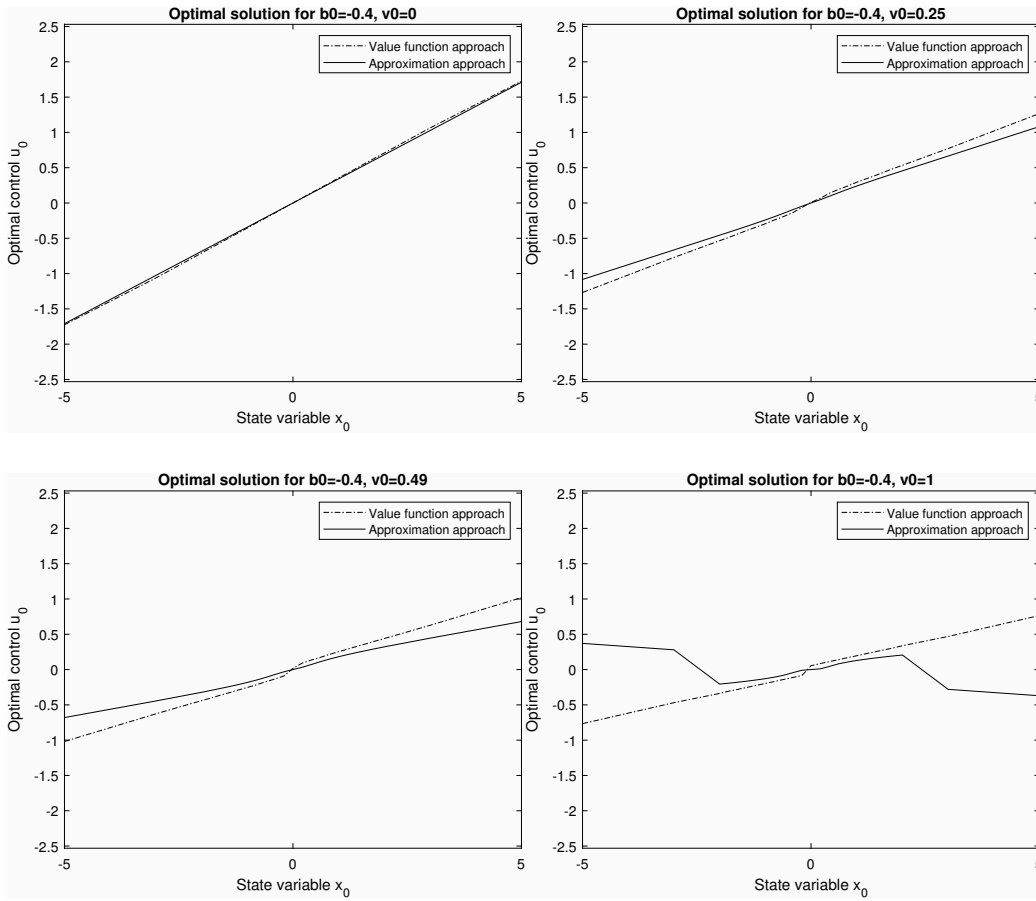


Figure 8:
Plot for $b_0 = -1.0$

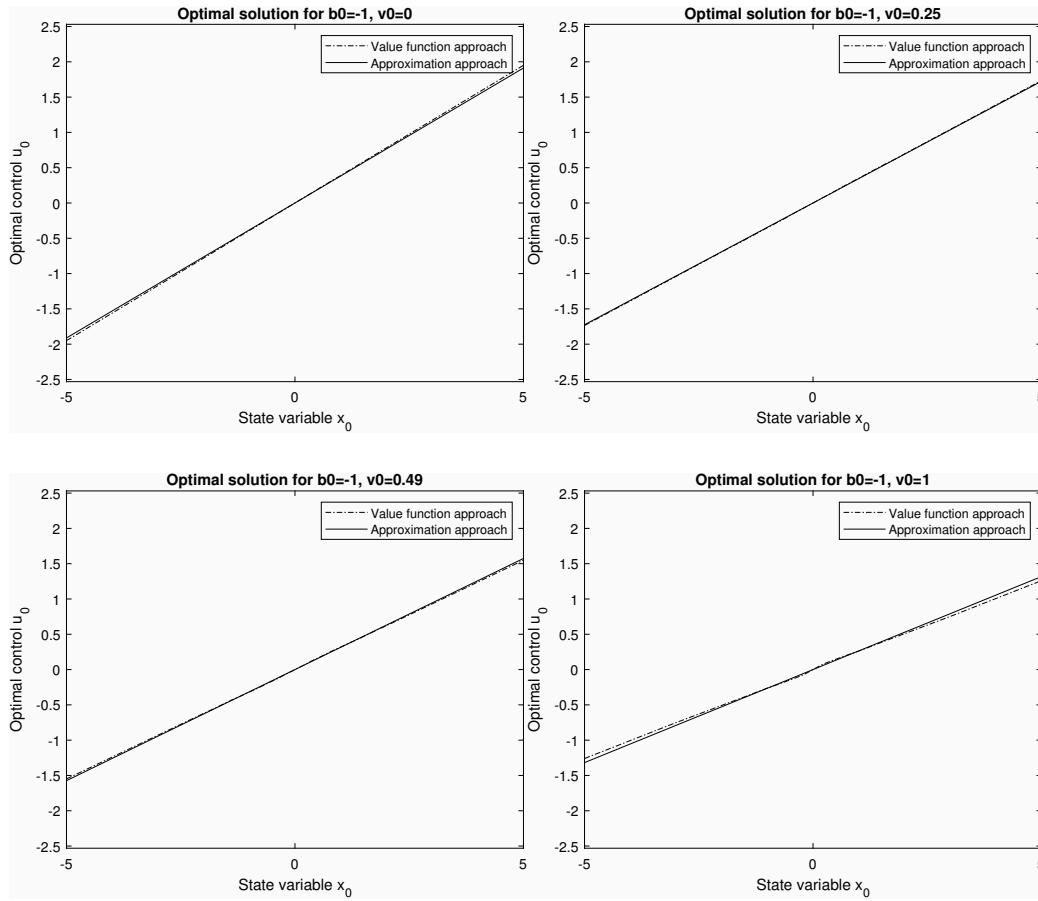


Figure 9:
Plot for $b_0 = -2.0$

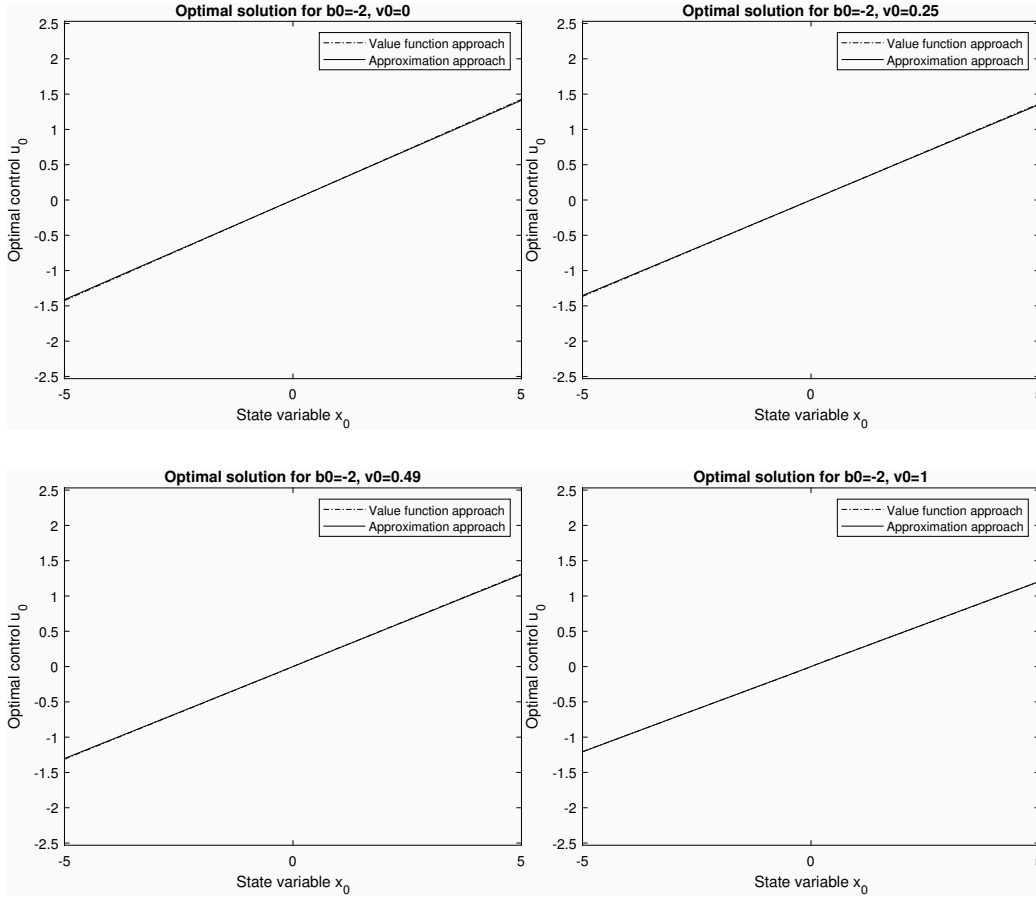
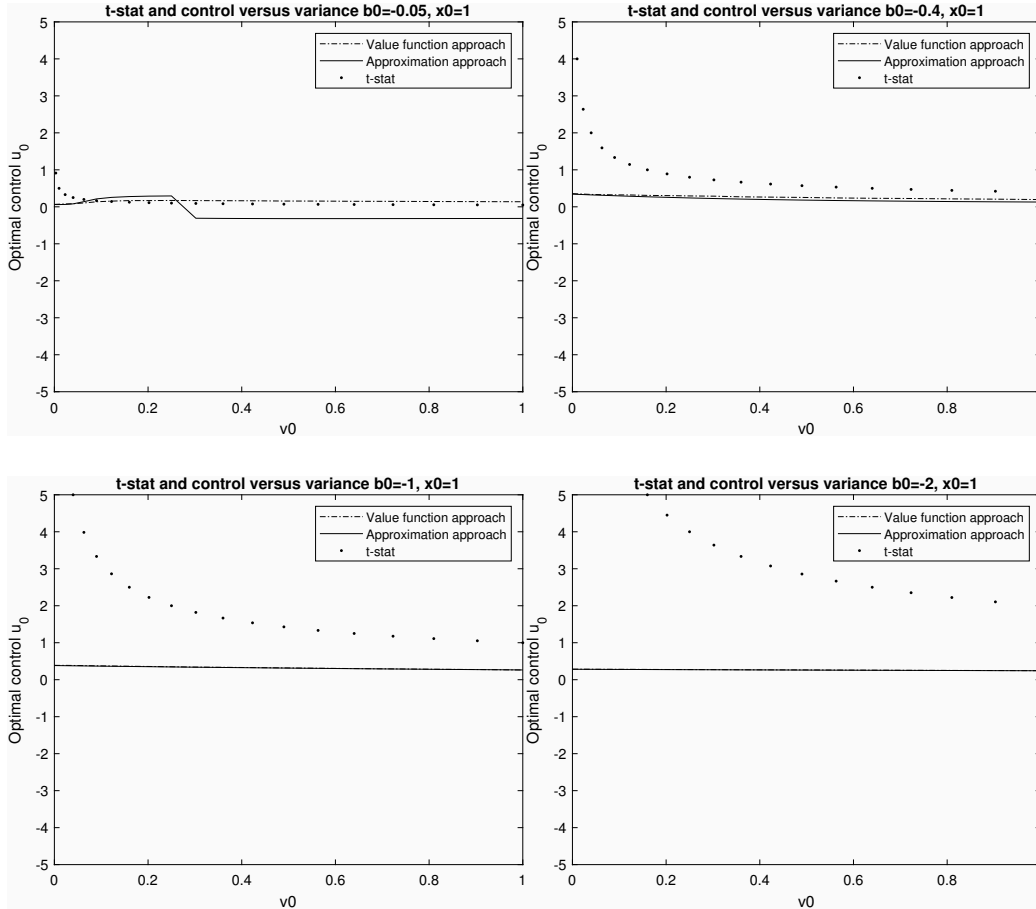


Figure 10 uses the same four values of b_0 to compare the two methods at various variances, when the initial state is $x_0 = 1$ and the system variance is $q = 0.1$. As in Figure 10 the difference is more noticeable when the t -statistics drops below 1.

Figure 10: Comparison Control versus Variance and t -statistics



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 Figure (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.

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 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 be 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., and Jullien, B. (1991). Optimal learning by experimentation. *Review of Economic Studies*, 58:621–654.
- Amman, H. and 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 Amman, H. M., Kendrick, D. A., and Rust, J., editors, *Handbook of Computational Economics*, volume 13 of *Handbook in Economics*, pages 579–618. North-Holland Publishers (Elsevier), Amsterdam, the Netherlands.
- Amman, H. M., Kendrick, D. A., and Tucci, M. P. (2018). Approximating the value function for optimal experimentation. *Macroeconomic Dynamics*.
- Amman, H. M. and Neudecker, H. (1997). Numerical solution methods of the algebraic matrix riccati equation. *Journal of Economic Dynamics and Control*, 21:363–370.
- Bar-Shalom, Y. and 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. and 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 University Press, Princeton, New Jersey.
- Bertsekas, D. P. (1976). *Dynamic Programming and Stochastic Control*, volume 125 of *Mathematics in Science and Engineering*. Academic Press, New York.
- Bolton, P. and Harris, C. (1999). Strategic experimentation. *Econometrica*, 67(2):349–374.
- Buera, F. J., Monge-Naranjo, A., and Primiceri, G. E. (2011). Learning the wealth of nations. *Econometrica*, 79(1):1–45.

- Coenen, G., Levin, A., and 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. (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.
- Cosimano, T. F. and Gapen, M. T. (2005a). Program notes for optimal experimentation and the perturbation method in the neighborhood of the augmented linear regulator problem. Working paper, Department of Finance, University of Notre Dame, Notre Dame, Indiana, USA.
- Cosimano, T. F. and Gapen, M. T. (2005b). Recursive methods of dynamic linear economics and optimal experimentation using the perturbation method. Working paper, Department of Finance, University of Notre Dame, Notre Dame, Indiana, USA.
- Easley, D. and Kiefer, N. M. (1988). Controlling a stochastic process with unknown parameters. *Econometrica*, 56:1045–1064.
- Hansen, L. P. and Sargent, T. J. (2007). *Robustness*. Princeton University Press, Princeton, NJ.
- Kendrick, D. A. (1981). *Stochastic control for economic models*. McGraw-Hill Book Company, New York, New York, USA. Second Edition, 2002.
- Kiefer, N. (1989). A value function arising in the economics of information. *Journal of Economic Dynamics and Control*, 13:201–223.
- Kiefer, N. and Nyarko, Y. (1989). Optimal control of an unknown linear process with learning. *International Economic Review*, 30:571–586.
- Levin, A., Wieland, V., and 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. and 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. and 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.
- Tucci, M. P. (2004). *The Rational Expectation Hypothesis, Time-varying Parameters and Adaptive Control*. Springer, Dordrecht, the Netherlands.
- Tucci, M. P., Kendrick, D. A., and 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, the Netherlands.