Macroeconomic Dynamics, **24**, 2020, 1073–1086. Printed in the United States of America.
doi:10.1017/S1365100518000664

APPROXIMATING THE VALUE FUNCTION FOR OPTIMAL EXPERIMENTATION

Hans M. Amman
University of Amsterdam

David A. Kendrick
University of Texas

Marco P. Tucci
University of Siena

In the economics literature, there are two dominant approaches for solving models with optimal experimentation (also called active learning). The first approach is based on the value function and the second on an approximation method. In principle the value function approach is the preferred method. However, it suffers from the *curse of dimensionality* and is only applicable to small problems with a limited number of policy variables. The approximation method allows for a computationally larger class of models, but may produce results that deviate from the optimal solution. Our simulations indicate that when the effects of learning are limited, the differences may be small. However, when there is sufficient scope for learning, the value function solution seems more aggressive in the use of the policy variable.

Keywords: Optimal Experimentation, Value Function, Approximation Method, Adaptive Control, Active Learning, Time-Varying Parameters, Numerical Experiments

1. INTRODUCTION

In the last decade, a large literature has emerged in economics on the subject of optimal or strategic experimentation also referred to as active learning. 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), and Moscarini and Smith (2001) and applications like Buera et al. (2011).

There are two prevailing methods for solving a general class of models with optimal experimentation. The first method is based on the value function approach

We would like to thank Volker Wieland for providing us with his software used in this paper. Furthermore, in writing this paper we have benefited greatly from the discussions we had with Tom Cosimano and Volker Wieland and the feedback from an anonymous referee. Address correspondence to: Hans M. Amman, Faculty of Economics and Business, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands; e-mail: amman@uva.nl, hans.amman@gmail.com. Mobile: +31651532162.

© 2018 Cambridge University Press 1469-8056/18

https://doi.org/10.1017/S1365100518000664 Published online by Cambridge University Press

and the second on an approximation method. The value function approach 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 (2000*a*,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 is 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*, 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), Cosimano and Gapen (2005*a*,b), Kendrick (1981), and Hansen and Sargent (2007) use approaches that are applied in the neighborhood of the linear regulator problems.[1,2] 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. 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 Section 4 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.

In this paper, we focus on comparing the policy function obtained by Beck and Wieland (2002), through the value function, to the one obtained through approximation methods. Note that the model used by Beck and Wieland (2002) closely resembles the model and framework presented in MacRae (1972, 1975), which has served as a benchmark in much of the literature.[3] In doing the comparison, we have derived an analytic form of the cost-to-go function for the approximation approach following the method outlined in the Amman and Kendrick (1995) paper and the extension of these results in Tucci et al. (2010).[4] After describing both of these methods in Section 3, we will present in Section 4 a comparison of the policy function results obtained through the value function approach and the approximation approach.

https://doi.org/10.1017/S1365100518000664 Published online by Cambridge University Press

2. A GENERAL FRAMEWORK

As the MacRae model (1972, 1975) is used as a benchmark in Beck and Wieland (2002), from here on we will use the abbreviation *MBW* (i.e. MacRae–Beck–Wieland) to refer to both sets of results in the literature. The general framework, Kendrick (1981), of this model starts with a linear difference equation for a finite time horizon

$$x_{t+1} = A_t(\theta_t)\, x_t + B_t(\theta_t)\, u_t + c_t(\theta_t) + \epsilon_t, \tag{1}$$

where $t \in [0, N-1]$ is the time index, $x_t \in \Re^{(n\times 1)}$ the state vector, $u_t \in \Re^{(m\times 1)}$ the control vector, $\epsilon_t \in \Re^{(n\times 1)}$ a vector of system noise terms, $A_t(\theta_t) \in \Re^{(n\times n)}$ the state vector coefficient matrix, $B_t(\theta_t) \in \Re^{(n\times m)}$ the control vector coefficient matrix, $c_t(\theta_t) \in \Re^{(n\times 1)}$ the exogenous coefficient vector, and $\theta_t \in \Re^{(s\times 1)}$ a vector containing the subset of the coefficients in $A_t(\theta_t)$, $B_t(\theta_t)$, and $c_t(\theta_t)$ that are stochastic. The measurement equation of the state x_t has the shape

$$y_t = H_t x_t + \xi_t, \tag{2}$$

where $y_t \in \Re^{(r\times 1)}$ is a measurement vector, $H_t \in \Re^{(r\times n)}$ a known measurement coefficient matrix, and $\xi_t \in \Re^{(r\times 1)}$ a vector of measurement noise terms. The time-varying parameter equation for θ_t is

$$\theta_{t+1} = D_t \theta_t + \eta_t, \tag{3}$$

where $D_t \in \Re^{(s\times s)}$ is a known matrix and $\eta_t \in \Re^{(s\times 1)}$ a vector of parameter noise terms.[5] The vector noise terms for the system, measurement, and parameter evolution equations are normally distributed

$$\epsilon_t \overset{i.i.d.}{\sim} N(0, \Sigma^{\epsilon\epsilon}) \qquad \xi_t \overset{i.i.d.}{\sim} N\left(0, \Sigma^{\xi\xi}\right) \qquad \eta_t \overset{i.i.d.}{\sim} N(0, \Sigma^{\eta\eta}) \tag{4}$$

and the covariance matrices $\Sigma^{\epsilon\epsilon}$, $\Sigma^{\xi\xi}$, and $\Sigma^{\eta\eta}$ are assumed to be known. The initial conditions for the state vector (1) and parameter vector (3) are also normally distributed

$$x_0 \sim N\left(x_0, \Sigma_0^{xx}\right) \qquad \theta_0 \sim N\left(\theta_0, \Sigma_0^{\theta\theta}\right) \tag{5}$$

with the covariance matrices $\Sigma_0^{xx} \in \Re^{(n\times n)}$ and $\Sigma_0^{\theta\theta} \in \Re^{(s\times s)}$. Due to the linearity in (1), this normality is carried over in time. As both $\{x_t, \theta_t\}$ are stochastic, their (posterior) estimates $\{\hat{x}_{t|t}, \hat{\theta}_{t|t}\}$ will be obtained through Bayesian updating using the Kalman filter. The same holds for the corresponding estimated covariance matrices $\{\hat{\Sigma}_{t|t}^{xx}, \hat{\Sigma}_{t|t}^{\theta\theta}\}$.

The criterion function may be written as

$$J = E\left\{\delta^N L_N(x_N) + \sum_{t=0}^{N-1} \delta^t L_t(x_t, u_t)\right\}, \tag{6}$$

where $J \in \Re$ is the criterion value, E the expectations operator, $\delta \in [0, 1>$ the discount factor, $L_N \in \Re$ the criterion function for the terminal period N, $x_N \in \Re^{(n\times 1)}$

https://doi.org/10.1017/S1365100518000664 Published online by Cambridge University Press

the state vector for the terminal period N, $L_t \in \Re$ the criterion function for period t, $x_t \in \Re^{(n\times 1)}$ the state vector for period t, and $u_t \in \Re^{(m\times 1)}$ the control vector for period t. The two terms on the right-hand side of equation (6) are defined as

$$L_N(x_N) = \frac{1}{2}(x_N - \tilde{x}_N)' W_N (x_N - \tilde{x}_N) \tag{7}$$

and

$$L_t(x_t, u_t) = \frac{1}{2}\big[(x_t - \tilde{x}_t)' W_t (x_t - \tilde{x}_t) + (x_t - \tilde{x}_t)' F_t (u_t - \tilde{u}_t) + (u_t - \tilde{u}_t)' \Lambda_t (u_t - \tilde{u}_t)\big], \tag{8}$$

where $\tilde{x}_N \in \Re^{(n\times 1)}$ is the desired state vector for terminal period N, $W_N \in \Re^{(n\times n)}$ the symmetric state variable penalty matrix for terminal period N, $\tilde{x}_t \in \Re^{(n\times 1)}$ the desired state vector for period t, $\tilde{u}_t \in \Re^{(m\times 1)}$ the desired control vector for period t, $W_t \in \Re^{(n\times n)}$ the symmetric state variable penalty matrix for period t, $F_t \in \Re^{(n\times m)}$ the penalty matrix on state-control variable deviations for period t, and $\Lambda_t \in \Re^{(m\times m)}$ the symmetric control variable penalty matrix for period t.

This completes the description of the general framework that will be applied in Section 4. In the next section, we describe briefly how the policy function can be derived for the approximation method.[6]

3. SOLUTION METHODS FOR THE GENERAL FRAMEWORK

The value function approach, though computational intensive, is quite straight forward, for example, integrate the value V function backward in time, starting at $N-1$

$$V(x_{N-1}, u_{N-1}) = \min_{u_{N-1}} \left\{ L_{N-1}(x_{N-1}, u_{N-1}) + \delta E\, V(x_N | \mathcal{P}_{N-1}) \right\} \tag{9}$$

subject to equation (1). The set $\mathcal{P}_{N-1}$ contains, under the normality assumption, the means and variance of the stochastic elements in equations (1)–(5). If we ignore measurement error, this means that

$$\mathcal{P}_{N-1} = \left\{ x_{N-1}, \Sigma^{xx}_{N-1}, \theta_{N-1}, \Sigma^{\theta\theta}_{N-1} \right\}.$$

Equation (9) is solved using discretization—see Judd (1998) or Kendrick et al. (2006a). Following Easley and Kiefer (1988) and Kiefer and Nyarko (1989), we can obtain, for $N \to \infty$, the optimal stationary policy, h_v, from the value function

$$u^*_v = h_v(x_0), \tag{10}$$

where u^*_v is the optimal response given the initial state x_0 in the value function approach.

https://doi.org/10.1017/S1365100518000664 Published online by Cambridge University Press

Now let's turn to the approximation approach. The approximation approach solves the cost-to-go function

$$J_N^* = \min_{\mathbf{u}_0} E\Big\{ \cdots \min_{\mathbf{u}_{N-2}} E\Big\{ \min_{\mathbf{u}_{N-1}} E\Big\{ C_N | \mathcal{Q}_{N-1} \Big\} | \mathcal{Q}_{N-2} \Big\} \cdots | \mathcal{Q}^0 \Big\} \tag{11}$$

subject to (1) with

$$C_{N-j} = L_N(x_N) + \sum_{t=j}^{N-1} L_t(x_t, u_t) \tag{12}$$

for $j = 0, \ldots, N-1$ and $\mathcal{Q}_j$, like $\mathcal{P}_j$, is defined as the means and covariances of the stochastic elements at time j. The essential difference between (9) and (11) is in the way $\mathcal{P}_j$ and $\mathcal{Q}_j$ are treated in the optimization. With the value function, the stochastic elements are fully integrated backward in time over the full probability space.

In the approximation approach, first an *certainty equivalence* solution is derived, Bertsekas (1976), for the problem in (1)–(8) using dynamic programming. Then, $\forall k \in [t, \ldots, N-1]$ the estimated stochastic elements

$$\left\{ \hat{x}_{k|k}, \hat{\Sigma}^{xx}_{k|k}, \hat{\theta}_{k|k}, \hat{\Sigma}^{\theta\theta}_{k|k} \right\}$$

are projected forward in time in the neighborhood of this certainty equivalence problem using Bayesian updating. Once a new set of estimates is obtained, the control set $\{u_k, \ldots, u_{N-1}\}$ is recomputed. These steps are repeated until convergence is reached. Through this forward projection, the combinatorial explosion in (9) is avoided and the resulting algorithm can be solved in polynomial time with $O(n^3)$.[7]

Based on equations (11) and (12) and following Tucci et al. (2010), the cost-to-go J_N can be divided into three terms, that is,

$$J_N = J_D + J_C + J_{P,} \tag{13}$$

J_D being the deterministic component, J_C the cautionary component, and J_P the probing component. The probing component is the component that represents the experimentation—active learning—part of J_N. The optimal is obtained by minimizing J_N[8]

$$u_a^* = \arg\min_u J_N\,(x_0, u), \tag{14}$$

which is similar to equation (10) and provides the possibility to (numerically) derive the policy function

$$u_a^* = h_a(x_0), \tag{15}$$

where u_a^* is the optimal response given the initial state x_0 in the approximation approach. In the next section, we will analyze the outcomes for both the policy functions (10) and (15).

https://doi.org/10.1017/S1365100518000664 Published online by Cambridge University Press

4. COMPARISON OF THE MBW POLICY FUNCTIONS

In order to make a comparison between the results based on the value function and those of the approximation method, we used Wieland's software and his parameter set.[9] We can write the MBW model, using the framework of Section 2, in its scalar form as follows:

Given the difference equation

$$x_{t+1} = \alpha x_t + b u_t + c_t \epsilon_t. \tag{16}$$

Find $\{u_0, \ldots, u_{N-1}\}$ to minimize

$$J = \frac{1}{2}\delta^N w_N (x_N - \tilde{x}_N)^2 + \frac{1}{2}\sum_{t=0}^{N-1} \delta^t \left\{ w_t (x_t - \tilde{x}_t)^2 + \lambda_t (u_t - \tilde{u}_t)_t^2 + \psi_t (x_t - \tilde{x}_t)(u_t - \tilde{u}_t) \right\}. \tag{17}$$

For comparison with the general framework in Section 2, this means that $A_t = \alpha = 1$, $c_t = 0$, $W_t = w_t = 1$, $\Lambda_t = \lambda_t = 0$, $F_t = \psi_t = 0$, $\tilde{x}_t = 0$, $\tilde{u}_t = 0 \ \ \forall\, t$, and $\Sigma^{\epsilon\epsilon} = \sigma_\epsilon^2 = 1$. The initial estimate of the stochastic parameter enters only the B matrix, so $B(\hat{\theta}_{0|0}) = \hat{\theta}_{0|0} = \hat{b} = -0.3$, with estimated variance $\hat{\Sigma}_{0|0}^{\theta\theta} = \hat{\sigma}_b^2 = 0.25$. There is no measurement error, which means $H = 1$ and $\Sigma^{\xi\xi} = \sigma_\xi^2 = 0$ and no parameter evolution, hence $D = 1$ and $\Sigma^{\eta\eta} = \sigma_\eta^2 = 0$ and as a consequence $\forall t$ $\hat{\theta}_{t+1|t} = \hat{\theta}_{t|t} = \hat{b}$. The initial estimate of the state $\hat{x}_{0|0} = \hat{x}_0 = x_0$ and its variance $\hat{\Sigma}_{0|0}^{xx} = \hat{\sigma}_x^2 = 0$, the discount factor $\delta = 0.95$, which completes the parameter set. Following Beck and Wieland (2002) we will focus on the stationary outcome, hence $N \to \infty$.

Since the MBW model has a single control variable and a single state variable, it is possible to derive an analytic form of the cost-to-go for the approximation approach. This is based on earlier work done in Amman and Kendrick (1995) and Tucci et al. (2010), when we analyzed the question whether or not the cost-to-go function was characterized by non-convexities. The three components in equation (13) are[10]

$$J_N = J_P + J_C + J_D, \tag{18}$$

$$J_D = \frac{1}{2}\,\delta\left(x_0 + \hat{b}u\right)^2, \tag{19}$$

$$J_C = \frac{1}{2}\,\frac{\delta\left[\hat{\sigma}_b^2(1-\delta)u^2 + \sigma_\epsilon^2\right]}{(1-\delta)}, \tag{20}$$

$$J_P = \frac{1}{2}\,\frac{\hat{\sigma}_b^2\delta^2\left[\left(\hat{\sigma}_b^2 - 1\right)u^2 + \sigma_\epsilon^2\right]\left[x_0 + 2\hat{b}u\right]^2}{\hat{b}^2(1-\delta)\left(\hat{\sigma}_b^2 u^2 + \sigma_\epsilon^2\right)}. \tag{21}$$

By minimizing $J_N(u, x_0)$, we get the policy function $u_a^* = h_a(x_0)$ from (14) as the approximated optimal response, that we can compare with value function results, (10), $u_v^* = h_v(x_0)$, derived through numerical integration. For our simulations, we used a grid for $x_0 \in [-5, 5]$ and a line search for $u \in [-20, 20]$, analyzing the

https://doi.org/10.1017/S1365100518000664 Published online by Cambridge University Press

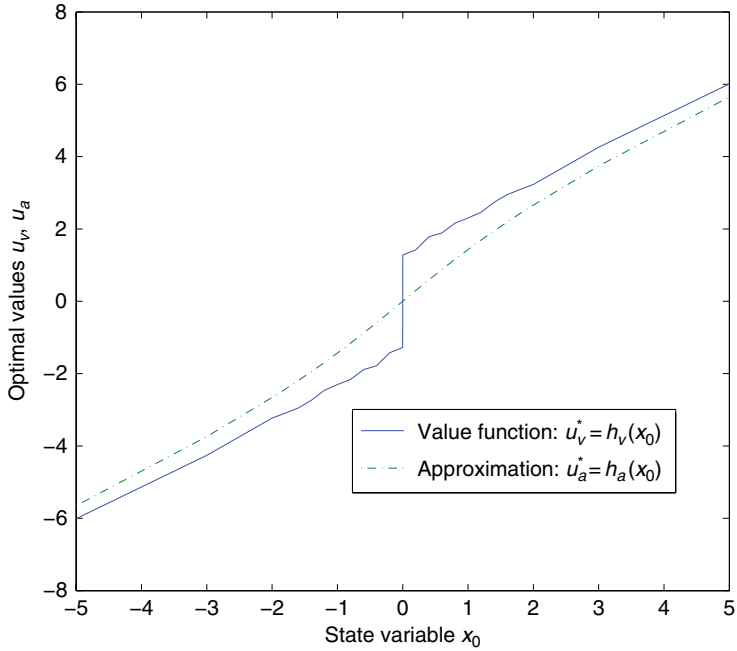


FIGURE 1. Baseline. $\delta = 0.95$, $\hat{b} = -0.30$, $\hat{\sigma}_b^2 = 0.25$, $\sigma_\epsilon^2 = 1.0$.

differences between the two approaches. We start with the baseline scenario using the above parameters. The results are shown in Figure 1.

With this baseline parameter set, a high discount rate and high variances, the value function solution has more aggressive control values than those in the approximation approach. This was to be expected as the value function, in principle, covers the whole solution space in comparison to the local approximation approach.[11] Please notice that for the Beck–Wieland (BW) model J_P is 0, when $x_0 = u = 0$. This stems from the fact that this is an equilibrium point where no gain is to be expected from experimentation as captured by the probing term J_P. In a more general setup with a nonzero constant term, nonzero desired path for the state and nonzero penalty on the controls. The probing component J_P will be > 0 even when $x_0 = u = 0$. See equation (11) in Tucci et al. (2010).

This is highlighted in Figure 2, where we plotted the three components of the cost-to-go for $x_0 = 0$. Especially for a sufficiently small $\varepsilon > 0$, $\{0 < |x_0| < \varepsilon\}$, the neighborhood of the stationary solution, the cautionary component J_C dominates the cost-to-go in the approximation, and the difference between the two methods are the largest.

As shown in Figure 3, in the "tails" of the plot of Figure 1, $\{0 < |x_0 - 5| < \varepsilon\}$, the experimentation, component J_P is dominant and the two methods tend to converge. Actually, the cost-to-go is non-convex and has a local minimum at $u_a = -4.579$.

https://doi.org/10.1017/S1365100518000664 Published online by Cambridge University Press

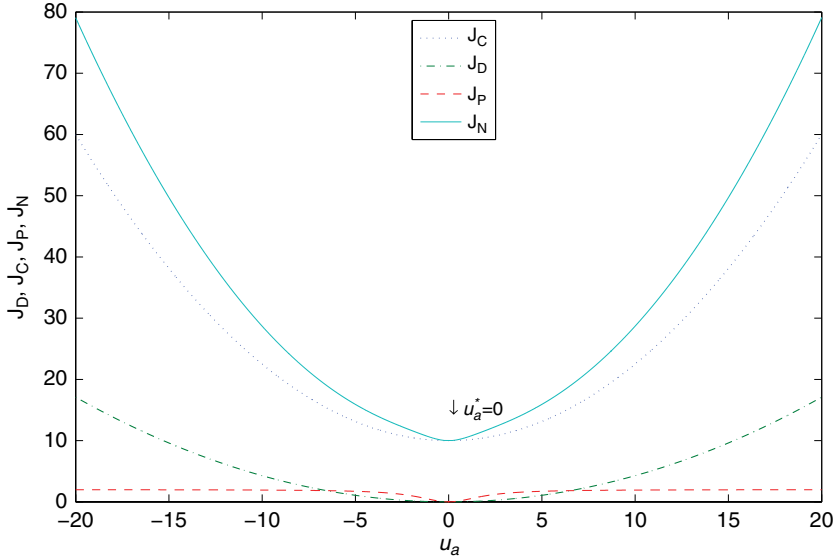


FIGURE 2. Cost-to-go. $\delta = 0.95$, $\hat{b} = -0.30$, $\hat{\sigma}_b^2 = 0.25$, $\sigma_\epsilon^2 = 1.0$, $x_0 = 0$.

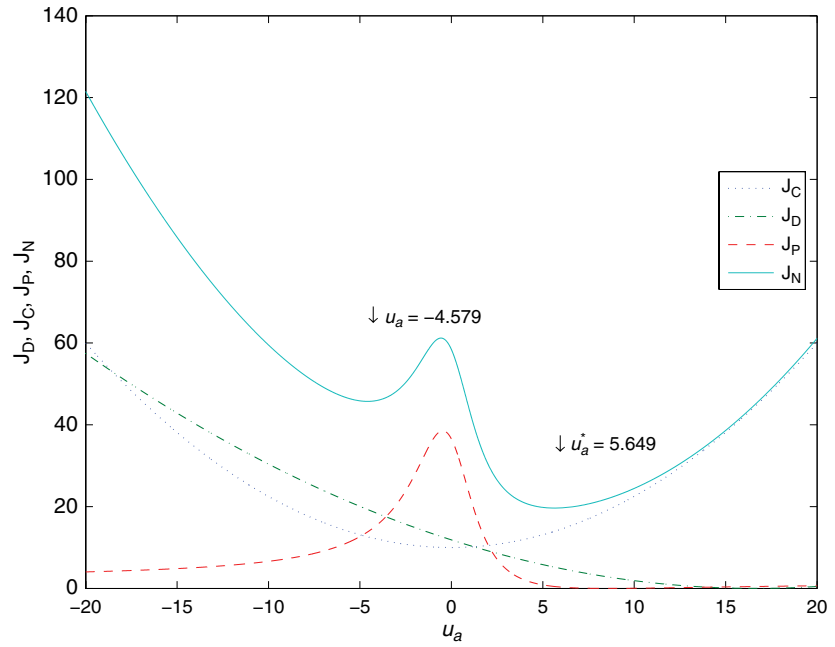


FIGURE 3. Cost-to-go. $\delta = 0.95$, $\hat{b} = -0.30$, $\hat{\sigma}_b^2 = 0.25$, $\sigma_\epsilon^2 = 1.0$, $x_0 = 5$.

https://doi.org/10.1017/S1365100518000664 Published online by Cambridge University Press

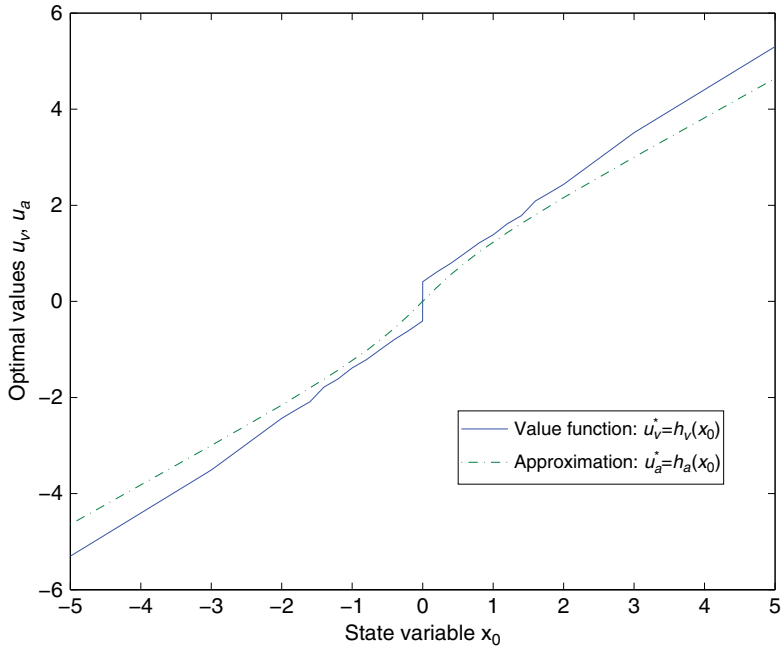


FIGURE 4. Lower discount factor δ. $\delta = 0.80$, $\hat{b} = -0.30$, $\hat{\sigma}_\theta^2 = 0.25$, $\sigma_\epsilon^2 = 1.0$.

On a more fundamental level, you see the "local nature" of the approximation method reflected in Figure 3. As it searches around the certainty equivalence path, in this case the equilibrium state, the experimentation is centered around $u = 0$.

To get more insight into the nature of the differences, we lowered the discount factor δ as shown in Figure 4. A lower δ means that the value of future experimentation—learning—has a lower value and leads to less experimentation. As a consequence, the solutions are closer than in the baseline scenario.

A similar effect takes place in Figures 5 and 6, in which we simulated what happens to the policy variable if we lower and increase the noise in system σ_ϵ^2. As a consequence, there is less or more need for experimentation to bring down the uncertainty in the system. Once again, the value function solution is more aggressive near the stationary point 0 than the approximation solution.

This is illustrated once more by Figures 7 and 8, where we lowered and increased the initial estimated variance of the parameter, $\hat{\sigma}_b^2$.

The overall picture is that when the BW parameter set is used, the approximation approach is a good alternative for the value function approach as long as the uncertainty in the model is low to modest.

Finally, we would make a comment on performance and feasibility. One of the reasons to use an approximation method is its feasibility. The approximation method can be solved in polynomial time, while the value function is non-polynomial problem due to its grid structure both in time and parameter space.

https://doi.org/10.1017/S1365100518000664 Published online by Cambridge University Press

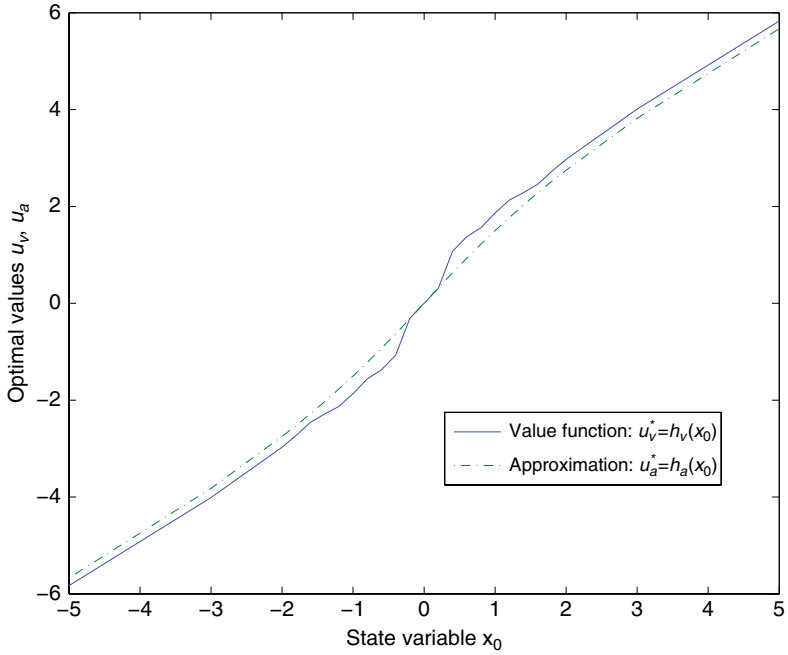


FIGURE 5. Lowering noise. $\delta = 0.95$, $\hat{b} = -0.30$, $\hat{\sigma}_b^2 = 0.25$, $\sigma_\epsilon^2 = 0.10$.

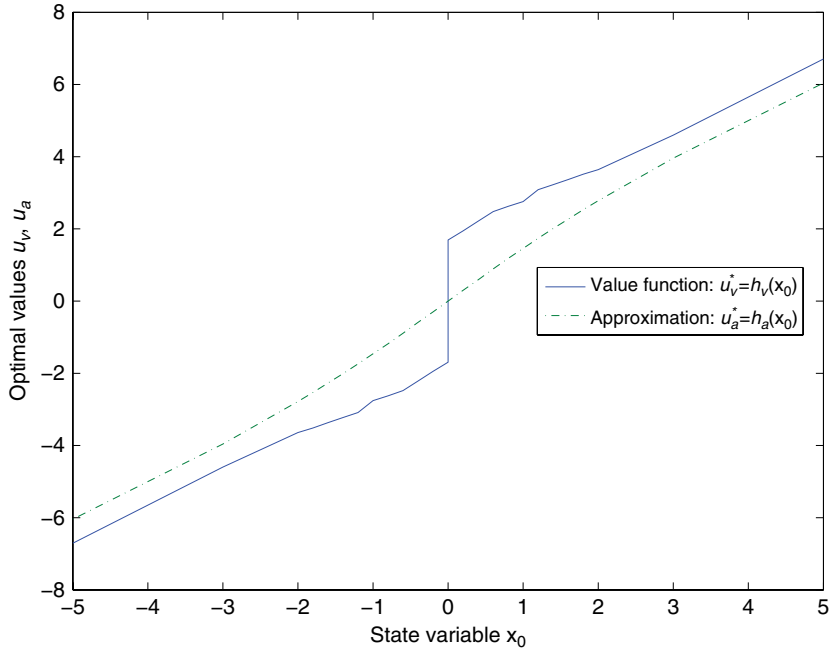


FIGURE 6. Increasing noise. $\delta = 0.95$, $\hat{b} = -0.30$, $\hat{\sigma}_b^2 = 0.25$, $\sigma_\epsilon^2 = 2.00$.

https://doi.org/10.1017/S1365100518000664 Published online by Cambridge University Press

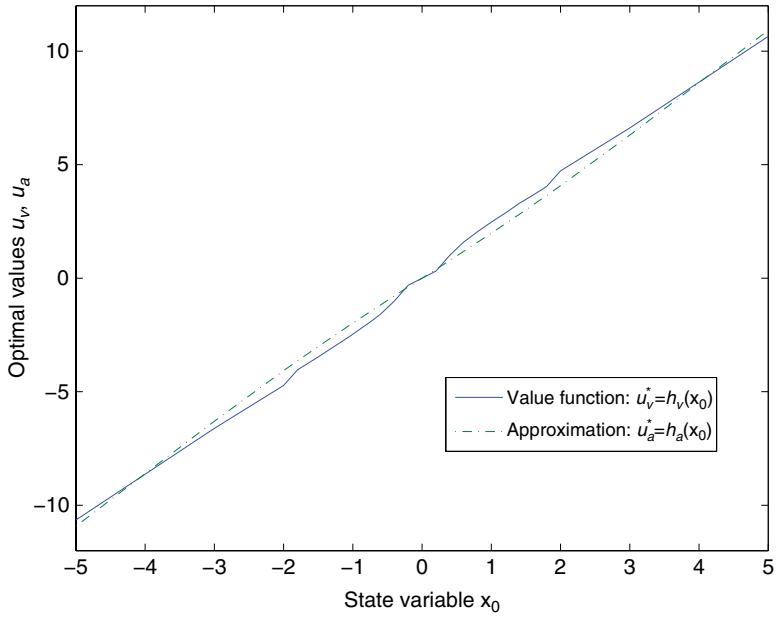


FIGURE 7. Lowering parameter uncertainty. $\delta = 0.95$, $\hat{b} = -0.30$, $\hat{\sigma}_b^2 = 0.04$, $\sigma_\epsilon^2 = 1.0$.

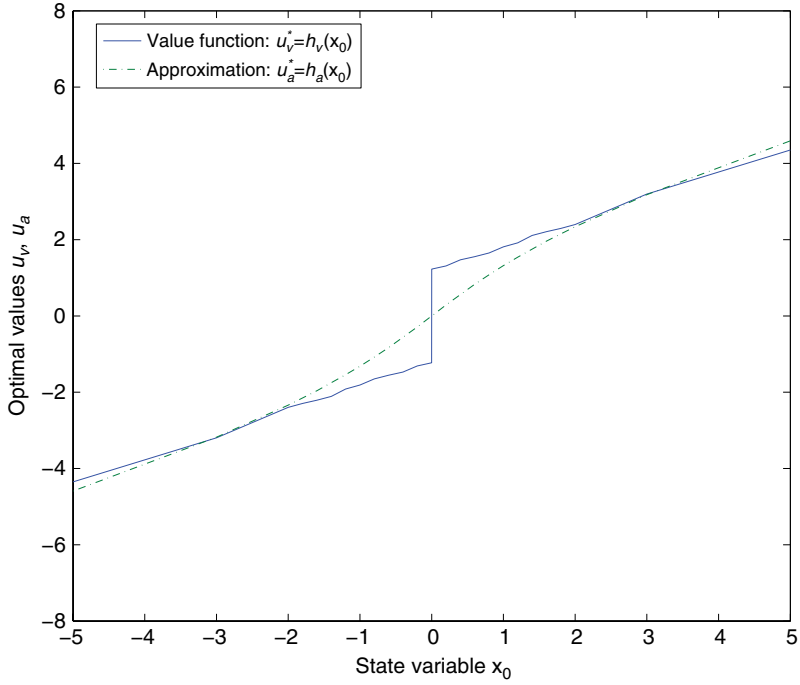


FIGURE 8. Increasing parameter uncertainty. $\delta = 0.95$, $\hat{b} = -0.30$, $\hat{\sigma}_b^2 = 0.49$, $\sigma_\epsilon^2 = 1.0$.

https://doi.org/10.1017/S1365100518000664 Published online by Cambridge University Press

Algorithmic efficiency and implementation is an art in its own,[12] which we would not like to discuss here. However, to give a crude indication from a practical perspective, it requires about 0.1 second to solve the approximation method in Figure 1 in *Matlab* and about 90 seconds to solve the value function approach with *Fortran*.

It may be useful to emphasize that the value function solution heavily depends on the grid used to solve the integral in the value function.[13] Hence, by reducing the grid to a more specific parameter set, the solution time can be reduced. On the other hand, the approximate solution is computed directly from the analytical formulae in the text.

5. CONCLUSIONS

Although the above results are in line with our intuition, we need to be modest. In this short paper, we provided some evidence that for the MBW model, the approximation method produces a solution close enough to the value function approach. Moreover, we have been able to identify some elements of the model specifications which affect the difference between the value function and the approximation solutions. Our simulations indicate that when the effects of learning are limited, the differences are small.

This is by no means a proof that approximation methods will provide results that are close to those obtainable with value function methods when a different parameter set is used or applied to larger models. The verdict is still out on this issue and is likely to stay out in the foreseeable future given the computational inefficiency of the value function method and the complexity of the issue.

NOTES

1. 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).

2. There are difference between the approximation approaches in Cosimano (2008) and in Kendrick (1981), which we will not discuss in detail here. Throughout the paper we will use the approach in Kendrick (1981).

3. See Amman and Kendrick (2008).

4. An extension of our approach, including rational expectations, was presented in Amman and Kendrick (1999).

5. For more general forms of equation (3), including the *return to normality* model, see Tucci (2004, p. 17).

6. The general framework in this section is coded in two software packages *DualPC*, Amman and Kendrick (1990), and more recently in *Duali*, Kendrick et al. (2006b). The framework may also be extended for the situation with rational expectations—cf. Amman and Kendrick (2003) and Kendrick et al. (2014).

7. The dimension of the state x_t, n, is usually the largest dimension and matrix operations of the order n^3 define the degree of complexity in the algorithm.

8. Once again we would like to stress that, quite frequently, equation (13) may have multiple minima as in Amman and Kendrick (1995) and Kendrick (1978).

https://doi.org/10.1017/S1365100518000664 Published online by Cambridge University Press

9. The software is available through the web site http://www.volkerwieland.com. A detailed discussion on how to compute the value function is found in Kendrick et al. (2006a).

10. The derivations of equations (19)–(21) are rather lengthy and are similar to those in Amman and Tucci (2017). The derivations are different from the one presented in Amman and Kendrick (1995) and Tucci et al. (2010), where we investigate the situation for $N \to \infty$ with $\delta \in [0, 1>$.

11. Note that, in Figure 1, for a sufficiently small $\varepsilon > 0$, $\{0 < |x_0| < \varepsilon\}$, the solution for u_v^* is somewhat jagged, which is caused by the discretization and has no economic interpretation. The same occurs for δ close to 1, causing numerical instability when integrating the value function. At this stage it is unclear how dependent are BW results on the grid for $\hat{b}$, $\hat{\sigma}_b^2$, and x_0 used in the original paper.

12. Cf. Amman (1989).

13. Beck and Wieland (2002), equation (7).

REFERENCES

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

Amman, H. M. (1989) Nonlinear control simulation on a vector machine. *Parallel Computing* 10, 123–127.

Amman, H. M. (1996) Numerical methods for linear-quadratic models. In: H. M. Amman, D. A. Kendrick and J. Rust (eds.), *Handbook of Computational Economics*. Handbook in Economics, Volume 13, pp. 579–618. Amsterdam, The Netherlands: North-Holland Publishers (Elsevier).

Amman, H. M. and M. P. Tucci (2017) The Dual Approach in an Infinite Horizon Model. Quaderni del Dipartimento di Economia Politica 766, Università di Siena, Siena, Italy.

Amman, H. M. and D. A. Kendrick (1990) A User's Guide for DUAL: A Program for Quadratic-Linear Stochastic Control Problems. Technical Report T90-4, Center for Economic Research, University of Texas, Austin, Texas, USA.

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

Amman, H. M. and D. A. Kendrick (1999) Linear-quadratic optimization for models with rational expectations. *Macroeconomic Dynamics* 3, 534–543.

Amman, H. M. and D. A. Kendrick (2003) Mitigation of the Lucas critique with stochastic control methods. *Journal of Economic Dynamics and Control* 27, 2035–2057.

Amman, H. M. and D. A. Kendrick (2008) Comparison of policy functions from the optimal learning and adaptive control framework. Discussion paper no. 08-19, Tjalling Koopman Institute, Utrecht School of Economics, Utrecht University, Utrecht, The Netherlands.

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

Beck, G. and V. Wieland (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*. Mathematics in Science and Engineering, Volume 125. New York: Academic Press.

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

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

Coenen, G., A. Levin and V. Wieland (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 M. T. Gapen (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.

https://doi.org/10.1017/S1365100518000664 Published online by Cambridge University Press

Cosimano, T. F. and M. T. Gapen (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 N. M. Kiefer (1988) Controlling a stochastic process with unknown parameters. *Econometrica* 56, 1045–1064.

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

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

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

Kendrick, D. A. (1981) *Stochastic Control for Economic Models*. New York, NY, USA: McGraw-Hill Book Company. Second Edition, 2002.

Kendrick, D. A., H. M. Amman and M. P. Tucci (2014) Learning about learning in dynamic economic models. In: K. Schmedders and K. Judd (eds.), *Handbook of Computational Economics*. Handbooks in Economics, Volume 3, Chapter 1, pp. 1–35. North-Holland: Elsevier.

Kendrick, D. A., P. R. Mercado and H. M. Amman (2006a) *Computational Economics (Supplementary chapters on the value function are provided by authors)*. Princeton and Oxford: Princeton University Press.

Kendrick, D. A., M. P. Tucci and H. M. Amman (2006b) DualI: Software for solving stochastic control problems. Working paper, Department of Economics, University of Texas, Austin, Texas, USA.

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

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

Levin, A., V. Wieland and J. C. Williams (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 L. Smith (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.

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*. Dordrecht, The Netherlands: Springer.

Tucci, M. P., D. A. Kendrick and H. M. Amman (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.

https://doi.org/10.1017/S1365100518000664 Published online by Cambridge University Press