Tjalling C. Koopmans Research Institute

Discussion Paper Series nr: 12-09

Conjectures on the policy function in the presence of optimal experimentation

Hans M. Amman
David A. Kendrick

Universiteit Utrecht

Utrecht School of Economics

Tjalling C. Koopmans Research Institute
Utrecht School of Economics
Utrecht University

Kriekenpitplein 21-22
3584 EC Utrecht
The Netherlands
telephone +31 30 253 9800
fax +31 30 253 7373
website www.koopmansinstitute.uu.nl

The Tjalling C. Koopmans Institute is the research institute and research school of Utrecht School of Economics.
It was founded in 2003, and named after Professor Tjalling C. Koopmans, Dutch-born Nobel Prize laureate in economics of 1975.

In the discussion papers series the Koopmans Institute publishes results of ongoing research for early dissemination of research results, and to enhance discussion with colleagues.

Please send any comments and suggestions on the Koopmans institute, or this series to J.M.vanDort@uu.nl

ontwerp voorblad: WRIK Utrecht

How to reach the authors

Please direct all correspondence to the first author.

Hans M. Amman
Utrecht School of Economics
Utrecht University
Heidelberglaan 8
3584 CS Utrecht
The Netherlands
h.m.amman@uu.nl

David A. Kendrick
Department of Economics
University of Texas
Austin, Texas 78712
USA
kendrick@austin.utexas.edu

Utrecht School of Economics
Tjalling C. Koopmans Research Institute
Discussion Paper Series 12-09

Conjectures on the policy function in the presence of optimal experimentation

Hans M. Amman[a]
David A. Kendrick[b]

[a]Utrecht School of Economics
Utrecht University

[b] Department of Economics
University of Texas

July 2012

Abstract

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 is more aggressive in the use of the policy variable.

Keywords: design of fiscal policy, optimal experimentation, stochastic optimization, time-varying parameters, numerical experiments.

JEL classification: C63, E61, E62

Acknowledgements

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, Marco Tucci and Volker Wieland.

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), 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 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 (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 is approach is 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 he 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 (2008) 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

[1] For consistency and clarity in the main text, we used the term approximation method in stead 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. Through out the paper we will use the approach in Kendrick(1981).

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 bench mark 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, Kendrick and Amman (2010). After describing both of these methods in Section 3, we will present in Section 4. of the paper a comparison of the policy function results obtained through the value function approach and the approximation approach.

2 A general framework

As the MacRae model (1972, 1975) is used as a bench mark in Beck and Wieland (2002), from here on we will use the abbreviation *MBW* 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

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

[3] See Amman and Kendrick (2008).

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, $\theta_t \in \Re^{(s\times 1)}$ 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, $H_t \in \Re^{(r\times n)}$ is a known measurement coefficient matrix $\xi_t \in \Re^{(r\times 1)}$ 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, $\eta_t \in \Re^{(s\times 1)}$ a vector of parameter-noise terms.[4] The a vector noise terms for the system, measurement and parameter evolution equations are normally distributed

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

and the covariances matrices $\Sigma^{\epsilon\epsilon}, \Sigma^{\xi\xi}, \Sigma^{\eta\eta}$ are assumed to be known. The initial conditions for the state vector (1) and parameter vector (3), that 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 covariances matrices $\Sigma_0^{xx} \in \Re^{(n\times n)}$, $\Sigma_0^{\theta\theta} \in \Re^{(s\times s)}$. Due to the linearity in (1), this normality is carried over in time. As both $\{\theta_t, x_t\}$ 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

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

[4] For more general forms of equation (3), including the *return to normality* model, see Tucci (2004) page 17.

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, $x_N \in \Re^{(n\times 1)}$ 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 $\tilde{x}_N \in \Re^{(n\times 1)}$ 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, $\Lambda_t \in \Re^{(m\times m)}$ 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[5]

3 Solution methods for the general framework

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

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

[5]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. (2006). The framework may also be extended for the situation with rational expectations, cf. Amman and Kendrick (2003) and Kendrick et. al (2013).

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

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

Equation (9) is solved using discretization, see Judd (1998) or Kendrick et. al (2006). 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.

Now lets turn to the approximation approach. The appraximation approach solves the cost-to-go function

$$J^*_N = \min_{\mathbf{u}_0} E\left\{ \ldots \min_{\mathbf{u}_{N-2}} E\left\{ \min_{\mathbf{u}_{N-1}} E\left\{C_N | \mathcal{Q}_{N-1}\right\} | \mathcal{Q}_{N-2} \right\} \cdots | \mathcal{Q}^0 \right\} \tag{11}$$

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, $\forall k \in [t, .., N-1]$ the estimated stochastic elements

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

are projected forward in time in the neighbourhood of this certainty equivalence problem using Bayesian updating. Once a new set of estimates is obtained, the control set $\{u_k, .., u_{N-1}\}$ 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 $O(n^3)$.[6]

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

$$J_N = J_D + J_C + J_P \tag{13}$$

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[7]

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

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).

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.[8] We can write the MBW model, using the framework of Section 2, in its scalar form as

[6] 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.

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

[8] 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 (2006).

Given the the difference equation

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

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

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

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 focuss on the stationary outcome, hence $N \to \infty$.

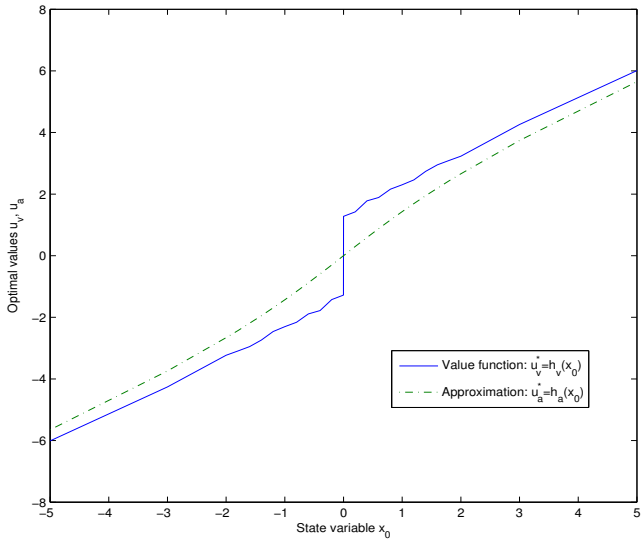
Since the MBW model has a single control variable and a single state variable it is possible to derive an analytic form of the cost-to-go for the approximation approach. This is based on earlier work done in Amman and Kendrick (1995) and Tucci et. al (2010), when we analyzed the question whether or not the cost-to-go function was characterized by non-convexities. The three components in equation (13) are[9]

[9]The derivation of equations (19)-(21) is rather lengthy. On request it can be provided by the corresponding author. The derivation is different from the one presented in Amman and Kendrick (1995) and Tucci et. al (2010), as we investigate the situation where $N \to \infty$ with $\delta \in [0, 1>$.

$$
\begin{aligned}
J_N &= J_P + J_C + J_D && (18)\\
J_D &= \frac{1}{2}\,\delta(x_0 + \hat{b}u)^2 && (19)\\
J_C &= \frac{1}{2}\,\frac{\delta[\hat{\sigma}_b^2(1-\delta)u^2 + \sigma_\epsilon^2]}{(1-\delta)} && (20)\\
J_P &= \frac{1}{2}\,\frac{\hat{\sigma}_b^2\delta^2[(\hat{\sigma}_b^2 - 1)u^2 + \sigma_\epsilon^2][x_0 + 2\hat{b}u]^2}{\hat{b}^2(1-\delta)(\hat{\sigma}_b^2 u^2 + \sigma_\epsilon^2)} && (21)
\end{aligned}
$$

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

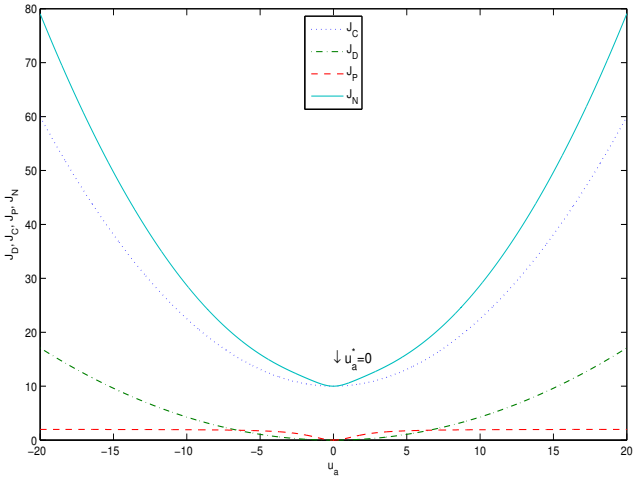
Figure 1: Baseline
$\delta = 0.95$, $\hat{b} = -0.30$, $\hat{\sigma}_b^2 = 0.25$, $\sigma_\epsilon^2 = 1.0$



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 to whole solution space in comparison to the local approximation approach.[10]

This is highlighted in Figure 2, where we plotted the three components of the 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.

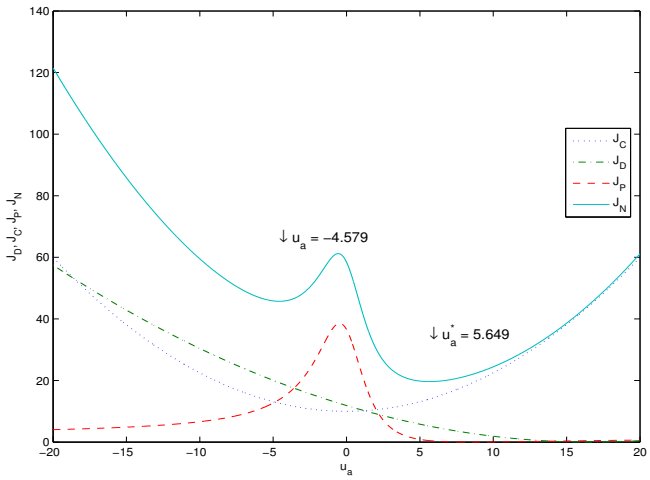
Figure 2: cost-to-go
$\delta = 0.95$, $\hat{b} = -0.30$, $\hat{\sigma}_b^2 = 0.25$, $\sigma_\epsilon^2 = 1.0$, $x_0 = 0$



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 convergence. Actual, the cost-to-go is non-convex and has a local minimum at $u_a = -4.579$.

[10] Note that in Figure 1, that 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.

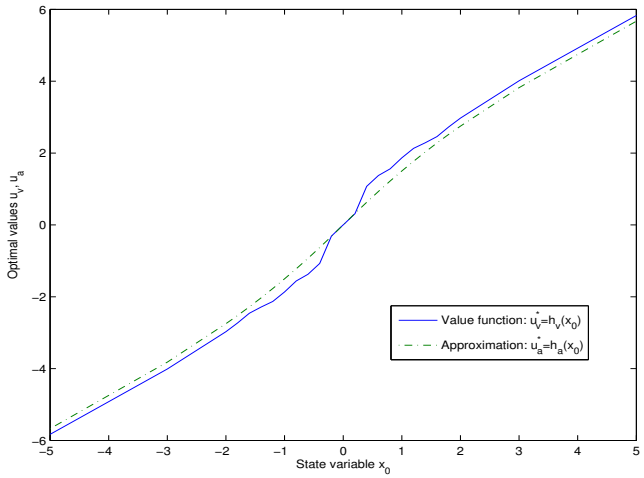
Figure 3: cost-to-go
$\delta = 0.95$, $\hat{b} = -0.30$, $\hat{\sigma}_b^2 = 0.25$, $\sigma_\epsilon^2 = 1.0$, $x_0 = 5$



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 equalibrium state, the experimentation is centered arond $u = 0$.

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.

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



A similar effect takes place in Figure 5 and 6, in which we simulated what happens to the policy variable if we lower and increase the noise in system σ^2_ϵ. 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.

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

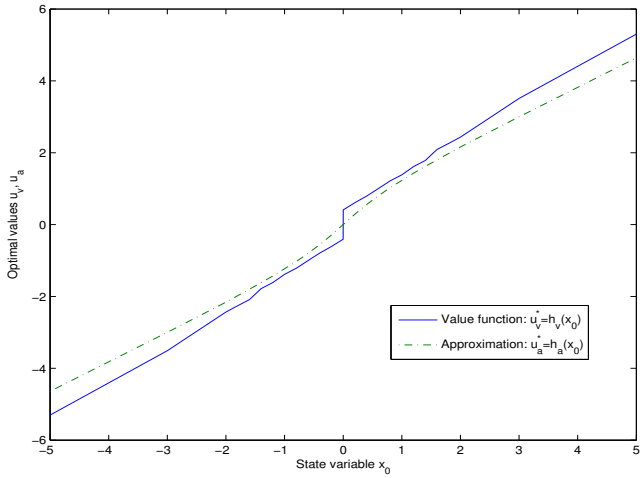
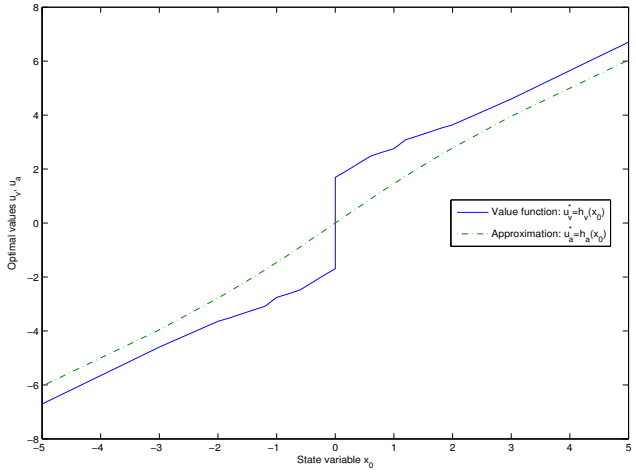


Figure 6: Increasing noise
$\delta = 0.95$, $\hat{b} = -0.30$, $\hat{\sigma}_b^2 = 0.25$, $\sigma_\epsilon^2 = 2.00$



This is illustrated once more by Figures 7 and 8 where we lowered and

increased the initial estimated variance of the parameter, $\hat{\sigma}_b^2$.

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

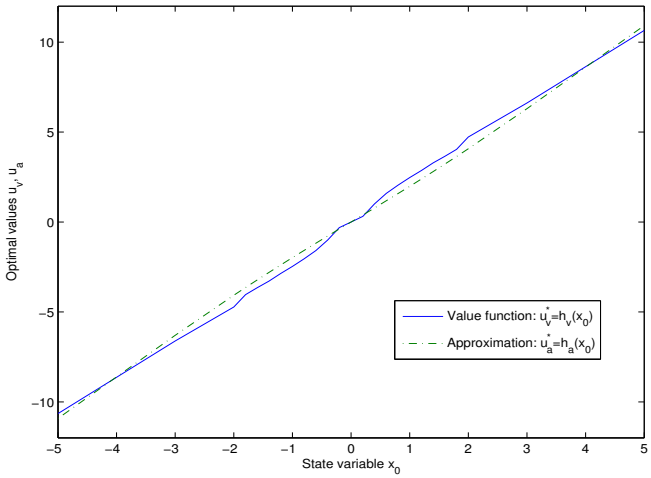
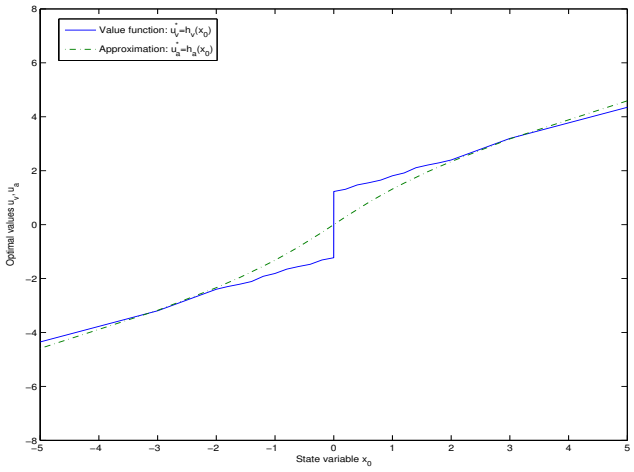


Figure 8: Increasing parameter uncertainty
$\delta = 0.95$, $\hat{b} = -0.30$, $\hat{\sigma}_b^2 = 0.49$, $\sigma_\epsilon^2 = 1.0$



The overall picture is that the approximation approach is a good alternative for the value function approach as long as the uncertainty in the model is low to modest.

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 MacRae-Beck-Wieland 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 in 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.

References

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

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*, Vol. 13 of *Handbook in Economics*, North-Holland Publishers (Elsevier), Amsterdam, the Netherlands, pp. 579–618.

Amman, H. M. and Kendrick, D. A.: 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 78712, USA.

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

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

Amman, H. M. and Kendrick, D. A.: 2008, Comparison of policy functions from the optimal learning and adaptive control framework, *Discussion Paper 08-19*, Tjalling Koopman Institute, Utrecht School of Economics, Utrecht University, Utrecht, the Netherlands.

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*, Vol. 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.: 2008, *Robustness*, Princeton University Press, Princeton, NJ.

Judd, K. L.: 1998, *Numerical methods in economics*, MIT Press, Cambridge, Massachussets.

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*, McGraw-Hill Book Company, New York, New York, USA. Second Edition, 2002. Available at url http://www.utexas.edu/cola/depts/economics/faculty/dak2?tab=127. **URL:** *http://www.utexas.edu/cola/depts/economics/faculty/dak2?tab=127*

Kendrick, D. A., Amman, H. M. and Tucci, M. P.: 2013, Learning about learning in dynamic economic models (forthcoming)., *in* K. L. Judd and K. Schmedders (eds), *Handbook on Computational Economics*, Vol. 3 of *Handbooks in Economics*, North-Holland Publishers, Amsterdam, the Netherlands.

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

Kendrick, D. A., Tucci, M. P. and Amman, H. M.: 2006, DualI: software for solving stochastic control problems, *Working paper*, Department of Economics, University of Texas, Austin, Texas 78712, 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 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.

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.