Volume Title: Annals of Economic and Social Measurement, Volume 6, number 5

Volume Author/Editor: NBER

Volume Publisher:

Volume URL: http://www.nber.org/books/aesm77-5

Publication Date: December 1977

Chapter Title: RESTRICTIONS ON THE CONTROL VECTOR IN ECONOMIC OPTIMIZATION PROBLEMS

Chapter Author: Roger Craine

Chapter URL: http://www.nber.org/chapters/c10539

Chapter pages in book: (p. 19 - 25)

Annals of Economic and Social Measurement, 6/5, 1978

RESTRICTIONS ON THE CONTROL VECTOR IN ECONOMIC OPTIMIZATION PROBLEMS

BY ROGER CRAINE*

This paper compares exact and "stochastic" smoothing restrictions on the control vector. It shows that restricting the k + 1st time difference of the control vector to zero is equivalent to restricting the control to a k degree polynomial function of time, and that the smoothing restriction can be relaxed by making the k + 1st difference stochastic. The restrictions are tested on a twelve-quarter optimization problem with the MPS model as a constraint. Solutions using exact restrictions are cheaper to compute, but are more sensitive to numeric problems.

INTRODUCTION

Restrictions on the control vector are a natural part of the problem specification in many economic applications, e.g., in macroeconomics a solution which yields a smooth instrument path is more politically acceptable than a policy with erratic changes. Restrictions should, and usually do, reduce the computations required to obtain a solution. McCarthy and Palash have shown that restricting the control path to a polynomial function of time reduces the dimensionality of a control problem the same way that restricting distributed lag weights to a polynomial reduces the dimensionality in an estimation problem. This paper extends their work in two directions: (1) I show that restricting the control vector to a k degree polynomial in time is equivalent to restricting the $k + 1^{st}$ time difference of the control vector to zero — this provides an easier and somewhat more intuitive way to impose exact polynomial restrictions; and (2) I allow the restriction on the $k + 1^{st}$ difference to be "stochastic"[1] which only forces the solution to lie in a band about the restriction.

Section I presents the two exact restriction procedures and shows that they are equivalent. Section II develops the stochastic restrictions and a measure of the marginal cost of the restriction. Section III reports some test results from applying the restrictions using the MPS model as a constraint.

*I wish to thank James Berry and Monica Friar of the Board of Governors of the Federal Reserve System for their help in computing these solutions, and the referees for their comments.

[1]This is equivalent to Shiller's distributed lag estimator.

I. Equivalence of Polynomial and Difference Restrictions on the Control

The technique McCarthy and Palash suggest is to restrict the time path of the control,[2] $u(t)$ to the k degree polynomial,

$$u(t) = \sum_{i=0}^{k} a_i t^i, \qquad t = 0, \ldots, T \tag{1}$$

where t is the t^{th} time period in the control horizon and the a_i are the parameters of the polynomial. When the degree of the polynomial, k, is less than the length of the control horizon T, only $k + 1$ parameters (the a_i) are needed to determine the $T + 1$ controls. In essence the controls are the a_i and the $u(t)$ are simply another endogenous variable in the model. As a result the dimensions in the control problem are reduced from $T + 1$ to $k + 1$ or by $T - k$.

An alternative way to impose the same restriction is to force the $k + 1^{st}$ difference of the control to be zero. For example, the $k + 1^{st}$ difference of the control is:

$$\Delta^{k+1} u(t) = \sum_{j=0}^{k+1} \binom{k+1}{j} (-1)^j u(t - j). \tag{2}$$

Substituting the k^{th} degree polynomial for $u(t - j)$ gives:

$$\Delta^{k+1} u(t) = \sum_{j=0}^{k+1} \binom{k+1}{j} (-1)^j \left[\sum_{i=0}^{k} a_i (t - j)^i \right] = 0. \tag{3}$$

That is, restricting the $k + 1^{st}$ time-difference of the control path to zero is equivalent to constraining the control path to lie on a k degree polynomial function of time. Setting equation (2) equal to zero and rearranging gives the entire control path as a function of the $k + 1$ initial conditions.

$$u(t) = \begin{cases} -\sum_{j=1}^{k+1} \binom{k+1}{j} (-1)^j u(t - j); & t = k + 1, \ldots, T \\ u^0(t) & : \ t = 0, \ldots, k. \end{cases} \tag{4}$$

The control problem is again reduced to $k + 1$ dimensions, except using (4) the parameters which must be found are the first $k + 1$ values of the control, $u^0(t)$.

The differencing procedure may have two trivial advantages over polynomial restrictions in computing openloop solutions to nonlinear

[2] The technique easily generalizes to a vector of controls with each control restricted to a polynomial which can be of different degree.

control problems using gradient techniques: (1) the "guesses" for the initial control values are likely to be closer to the iterated solution values than the guesses for the unituitive polynomial parameters,[3] and (2) since the difference equation is recursive in the parameters (lagged values) while the polynomial is not $k(k + 1)/2$ time periods of model simulation can be saved each time a gradient is computed.

II. "Stochastic" Restrictions

The major advantage to writing the restriction in difference form (which was pointed out by Shiller)[4] is that the restrictions can be easily transformed to a stochastic form by not forcing the restriction to hold exactly. Adding an error term to (4) gives

$$\Delta^{k+1}u(t) = e_t. \tag{5}$$

Shiller assumes the error is distributed with a zero mean and constant variance σ. In the deterministic control problem the error is not stochastic: instead it is the deviation from a smoothed path or an error from approximating the true minimizing control function with a low-order Taylor series expansion.

The stochastic restriction can be added to the original loss function $L(u)$ as:

$$L^R = L(u) + w\sum_{t=0}^{T}(\Delta^{k+1}u(t))^2. \tag{6}$$

The restricted loss function forces the $k + 1^{st}$ difference of the controls to lie in a band around zero: the larger the penalty weight, w, the smaller the band width. Each element of the control vector is still independent, however, since the restriction only concentrates the loss in the $k + 1$ parameter space of the difference equation but does not reduce it to exactly $k + 1$ dimensions.

The marginal cost of the constraint,

(7)
$$\frac{dL^R}{dw} = \sum_{t=0}^{T}\left[\frac{\partial L}{\partial u(t)}\frac{du(t)}{dw} + (\Delta^{k+1}u(t))^2 + 2w\frac{\partial\Delta u(t)}{\partial u(t)}\frac{du(t)}{dw}\right] \geq 0$$

shows the slope of the loss surface evaluated at a given weight $\bar{w}$.[5] The

[3] Schiller argues also that our priors are better about the desired smoothness of the process than they are about the degree of the polynomial restriction.

[4] Shiller proposed this technique to estimate distributed lag weights.

[5] The marginal cost of the constraint can be approximated numerically by

$$(\min L^R(\bar{w} + \Delta w) - \min L^R(\bar{w}))/\Delta w.$$

marginal cost should not be large unless there is an economic justification for not relaxing the constraint.

III. Comparison of Results

This section presents the results from test runs using exact difference or stochastic restrictions.

The restrictions were tested using a quadratic loss function that penalizes positive deviations in unemployment (u) from the "natural" rate of 4.8 percent and deviations in the inflation rate ($\dot{p}$ – the rate of change of the GNP deflator) from 2.5 percent over the twelve-quarter horizon 69I–71IV.

$$L = \sum_{t=69I}^{71IV} 2(u \geq 4.8)_t^2 + (\dot{p} - 2.5)_t^2. \tag{8}$$

The MPS (1970 version) quarterly econometric model was used as a constraint. The exogenous variables were set at their historical values[6] except for the control variable, the log of M1 which was chosen to minimize the loss function (8) subject to the exact or stochastic restrictions. We used a conjugate gradient algorithm[7] to determine the direction and a linear search to find the best step-size at each iteration in the optimization.

Table 1 shows the solution times[8] and iterations. Table 2 shows the

TABLE 1

Restriction	Iterations	CPU Time Min.	CPU Time Sec.	Loss Target	Loss Control
LM1 = history				113.77	
Δ^2LM1 = 0	16	2	56	90.10	
Δ^3LM1 = 0	12	2	03*	105.61	
$w(\Delta^2 LM1)^2$					Δ^2LM1
w = 100K	12	4	54	93.58	2.45
w = 75K	10	4	20	90.34	4.74
w = 50K	12	5	02	92.50	0.83
w = 25K	6	2	36	89.46	5.28
w = 1K	24	10	16	71.10	0.67
w = 100	13	5	39	70.26	0.21
$w(\Delta^3 LM1)^2$					Δ^3LM1
w = 75K	25	10	43**	92.83	0.61
w = 50K	10	4	12	91.37	7.01
w = 0	16	6	49***	70.14	

*Started from smoothed log M1 path; all others started from historical M1 path.
**Smaller maximum step-size and perturbation for derivative calculation.
***Started from solution path of run w = 100.

[6] No residuals were used.
[7] For example, see Kowalik and Osborne.
[8] All calculations were done on an IBM 370 model 168.

TABLE 2

GROWTH RATE OF M1

Restriction	$\Delta^2 LM1 = 0$	$75K(\Delta^2 LM1)^2$	$50K(\Delta^2 LM1)^2$	none
Time				
69I	-2.34	-4.18	-2.49	-20.67
69II	7.74	8.85	7.93	34.65
69III	.	8.60	7.53	32.17
69IV	.	8.58	7.15	13.92
70I	.	6.93	6.27	5.96
70II		5.38	5.29	2.03
70III		4.17	4.93	8.65
70IV		3.81	5.24	7.79
71I		5.11	5.64	1.88
71II		6.53	6.14	4.65
71III		7.38	6.53	0.32
71IV	7.74	7.81	6.15	3.14

control path for the growth rate of M1. The exact restrictions are that the second or third difference of log of M1 (Δ^2LM1, Δ^3LM1) is zero which implies a constant money growth rate or a constant rate of change of the money growth rate. The stochastic restrictions consist of a weight (w) on the second or third difference of log(M1) which penalizes derivations from a constant money growth rate or a constant rate of change in the money growth rate.

The results are close to what was anticipated. On average the exact restrictions took less CPU time since the computation of the gradient at each iteration took less time.[9] The lower dimension of the gradient did not reduce the number of iterations, however, as it would have in a linear-quadratic problem where the maximum iterations is given by the dimension of the control vector.[10]

Relaxing the constraint by lowering the weight (w) generally reduced the loss; furthermore, the marginal cost of the constraint was very low across a wide range of weights - weights between 100K and 25K gave very similar solutions - as long as the constrant was binding. This is encouraging because within this range the solution seems reasonable. Removing the constraint produces a large drop in the loss, but a politically unacceptable solution (column 4, table 2) and a solution which probably drives the model into an unreliable region.

The tables do not indicate dominance by either technique. In fact, there are a number of inconsistencies which again show that one must be careful when applying gradient techniques to large nonlinear (and non-

[9]There a fixed set-up time for each gradient calculation and since the problem is nonlinear the convergence is not uniform so that the CPU times varies between iterations.

[10]See Kowalik and Osborne, p. 40.

convex) problems. The algorithms may converge rapidly, but to a local minimum (e.g. w = 50K) or they may find a direction in which the loss function is very steep and converge rapidly to the proper minimum (e.g. w = 25K, w = 100). Consequently average performances are a better indicator than any single run.

The inconsistencies also point to some numerical problems. Relaxing the constraint by increasing the degree of the difference restriction (from two to three) resulted in an increase in the loss in two of three cases.[11] For the stochastic restrictions the solutions were very close to the comparable second difference runs. Since the marginal cost of the restriction is low for weights in this region (25Kto 100K) not much improvement--but, not a decrease in performance--should have been expected. In the case of the exact restriction the increase in the loss was substantially larger, and we only found a convergent solution after considerable experimentation[12] which is an indication of numerical problems.

The numerical accuracy of solutions were tested using a zero function.[13] We chose the solution paths from the exact second-order difference restriction (constant money growth) run as a target path and tested whether the different restricted experiments could "zero" this loss function (theoretically they could). The stochastic restrictions and the exact second difference restriction reached a minimum of 0.02 while the loss from the exact third difference restriction was around 2.2, confirming our suspicion of numerical difficulties.

IV. Conclusions

The results presented here suggest that constraints on the control path, either exact or stochastic, reduce the time required to compute a control solution; and what is probably more important, they constrain the solution to the region in which the model is a better approximation of the true economic structure. However, the results also indicate that neither technique can be applied mechanically with much hope of obtaining reasonable results. The exact restrictions appear to be more sensitive to numeric problems, but cheaper to compute.

University of California, Berkeley

[11] Since the original runs we tried a Davidon, Fletcher, Powell algorithm with the hope that information in the Hessian would eliminate some of the inconsistencies. Unfortunately the results essentially parallel the results in table 1.

[12] To find a solution we smoothed the starting path and successively reduced the maximum step-size and perturbations size for the gratient calculation until the algorithm converged, and still it converged to the relatively poor minimum. Cutting the step-size and perturbations further also gave explosive solutions.

[13] See Ando, Norman, and Palash for more detail.

BIBLIOGRAPHY

Almon, S., "The Distributed Lag Between Capital Appropriations and Expenditures." *Econometrica*, 33, no. 1 (January 1965): 178–196.

Ando, A.; Norman, A.; and Palash, C. "On the Application of Optimal Control to Large Scale Econometric Models." *Management Science*, in press.

Kolwalik, J., and Osborne, M. *Methods for Unconstrained Optimization Problems*. New York: Elsevier, 1968.

McCarthy, M., and Palash, C. "The Use of Almon- and Other Dummy-Variable Procedures to Increase the Efficiency of Maximization Algorithms in Economic Control." *Annals of Economic and Social Measurement*, 6/2, 1977, pp. 225–229.

Shiller, R. "A Distributed Lag Estimator Derived from Smoothness Priors." *Econometrica* 41, no. 4 (July 1973): 775–788.