Optimal Control of Large, Forward-looking Models

Efficient Solutions and Two Examples

Frederico S. Finan and Robert Tetlow*
Division of Research and Statistics
Board of Governors of the Federal Reserve System
Washington, DC 20551

Original: October 1999; Revised May 2000

Abstract

An optimal control tool is described that is particularly useful for computing rules of large-scale models where users might otherwise have difficulty determining the state vector *a priori* and where the inversion of large, sparse matrices can make working with matrix Ricatti equations numerically costly or infeasible. A small-scale demonstration is presented, as are data on performance with the Board of Governors' large-scale rational expectations macroeconometric model, FRB/US.

Keywords: optimal control, mathematical programs, rational expectations, simple rules.

JEL codes: C6, C8.

* Corresponding author: Tetlow at mail stop 61, telephone: (202) 452-2437, facsimile: (202) 452-5296, e-mail: Robert.J.Tetlow@frb.gov. We thank, without implication, Gary Anderson, Magnus Jonsson, Paul Klein, Brian Madigan and Peter von zur Muehlen for useful comments. All remaining errors are ours. The views expressed in this article are those of the authors only and are not necessarily shared by the Board of Governors or its staff.

1. Introduction

Rarely do economists compute optimal control rules for large-scale, linear rational expectations models. To do so obliges the researcher to overcome three important hurdles: (1) the model must be solved for its state-space representation; (2) those variables that are non-predetermined must be identified in advance so that the optimal rule can be solved in terms of those variables' costates instead of their states; and (3) iteration usually must be undertaken using large, sparse matrices. This article describes a technique that when used in conjunction with a particular method for finding saddle-point solutions of linear rational expectations models, overcomes these problems. Armed with these tools, a researcher needs only to be able to type a model into a front-end program and execute. Both the saddle-point solution algorithm--an implementation of the QR decomposition called AIM--and the optimal control program are downloadable.[1]

Two examples of the method are shown. The first example uses a simple sticky-price model of the business cycle. The second uses FRB/US, the Board of Governors' large-scale rational expectations macroeconometric model of the U.S. economy. With FRB/US, we show that simple rules are not always good approximations of optimal rules, contrary to what is sometimes claimed.

2. Method

We describe a method for computing the optimal control rule of large, linear rational expectations models; that is, models that can be written in the following form:

$$\sum_{i=-\tau}^{\theta} A_i \{X_{t+i} | \Omega_t\} = \varepsilon_t \tag{1}$$

where X_t is an *n-by-1* vector of endogenous variables, A_i, $i = -\tau, \ldots, \theta$ is a matrix of structural coefficients prior to control; τ is the maximum (finite) lag, and θ is the maximum lead, and ε_t is a vector of disturbance terms. The set Ω_t defines the information available to decision makers when they formulate expectations. The policymaker's loss function is quadratic:

1. The AIM algorithm and was developed in stages by Anderson and Moore (1985) and Anderson (2000). Versions of AIM are written in C, Matlab, Mathematica and Gauss. At present, the optimal control program is available only in Matlab. The Matlab version of AIM can be downloaded from the public website of the Federal Reserve Board at http://www.federalreserve.gov/pubs/oss/oss4/code.html while the optimal control program--called *LQcontrol.m*--can be downloaded from the FEDS working papers download site at http://www.federalreserve.gov/pubs/feds/1999/199951/199951code.zip

$$J_0 = \frac{1}{2}E_0\sum_{i=0}^{\infty}\beta^i(X'_{t+i}CX_{t+i} + X'_{t+i}GU_{t+i} + U'_{t+i}MU_{t+i}) \tag{2}$$

where U is a *k-by-1* vector of control variables, and C, G and M are conformable matrices of coefficients. The solution proceeds by forming the Lagrangian for this problem and solving for first-order conditions with respect to λ_{t+i}, X_{t+i}, U_{t+i} where λ_{t+i} is a vector of costate variables:[2]

$$\sum_{i=-\tau}^{\theta} A_i X_{t+i} + BU_{t+i} = \varepsilon_t \tag{3}$$

$$G'X_{t+i} + MU_{t+i} = -B'\lambda_{t+i} \tag{4}$$

$$CX_{t+i} + GU_{t+i} = A_0'\lambda_t + \sum_{i=1}^{\theta}\beta^{-i}A'_i\lambda_{t-i} + \sum_{i=-\tau}^{-1}\beta^i A'_i\lambda_{t-i} \tag{5}$$

where B is a column vector of parameters representing the influence of the control variable(s) on state variables, and A_0 is that column of A_i that is associated with contemporaneously dated variables.

Equations (3) are the structural equations of the model; equations (4) identify each control variable with a particular costate; and equations (5) describe the laws of motion for each costate. The matrix A_i is the matrix of structural coefficients for the model, prior to its extension to allow for control. Post-optimization, A_i is extended to include the preference parameters, β, C, G and M to form the (larger) matrix H:

$$\begin{array}{c}\left[\begin{array}{ccccccccccccccccc} A_{-\tau} & 0 & 0 & \dots & A_{-1} & 0 & 0 & \underset{\langle\text{n-by-n}\rangle}{A_0} & \underset{\langle\text{n-by-k}\rangle}{B} & \underset{\langle\text{n-by-n}\rangle}{0} & A_1 & 0 & 0 & \dots & A_\theta & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & \underset{\langle\text{k-by-n}\rangle}{G'} & \underset{\langle\text{k-by-k}\rangle}{M} & \underset{\langle\text{k-by-k}\rangle}{B'} & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & -\beta^{-\tau}A'_\theta & \dots & 0 & 0 & -\beta^{-1}A'_1 & C & G & -A'_0 & 0 & 0 & -\beta A'_{-1} & \dots & 0 & 0 & -\beta^\theta A_\tau' \end{array}\right] = H \\ \begin{array}{ccccc} H_{-\tau} & \quad H_{-1} & \quad H_0 & \quad H_1 & \quad H_\theta \end{array}\end{array} \tag{6}$$

Defining $Z_t = \begin{bmatrix} X_t & U_t & \lambda_t \end{bmatrix}'$ the system under control can be expressed as:

$$\begin{bmatrix} H_{-\tau} & \dots & H_0, H_1, \dots, H_\theta \end{bmatrix}\begin{bmatrix} Z_{t-\tau}, \dots & Z, Z_{t+1}, \dots, Z_{t+\theta+1} \end{bmatrix}' = \varepsilon_t \tag{7}$$

2. Using Lagrangians, instead of, say, dynamic programming to solve optimization problems is not new and has earned a rebirth in Chow (1997). The solution described here follows, in part, the commitment solution of the methods described by Levine and Currie (1987) and Backus and Driffill (1986), but is much quicker and more reliable. Extensions to the discretionary solutions would be straightforward to implement.

At this point, there are n equations and $(\theta + 1)n$ unknowns: θn equations or restrictions are needed to pin down a unique solution.

With the control problem now cast in the form of equations (7), one needs to compute the state-space representation of the structural model and the associated transition matrix, $\tilde{H}$:

$$\begin{bmatrix} Z_{-\tau+1} \\ \dots \\ Z_{\theta} \end{bmatrix} = \tilde{H} \begin{bmatrix} Z_{-\tau} \\ \dots \\ Z_{\theta-1} \end{bmatrix} \tag{8}$$

As we shall demonstrate, the matrix $\tilde{H}$ contains the same information as H but rearranged to be conformable with the state-space representation of the (costate-augmented) structural model. It would be a simple task to compute equations (8) were it not for the problem that the lead matrix H_{θ} in (6) is often singular. Solving this problem be done using any of several algebraic decompositions; we find the AIM algorithm of Anderson and Moore (1985) and Anderson (2000) to be particularly useful. A complete depiction of the intricacies of the AIM algorithm requires more space than we can devote here. Accordingly, we limit ourselves to providing an illustrative sketch. See Anderson (2000) for details.

The solution involves using equations (7) to: (i) find a state-space representation; (ii) compute the eigenvalues and the invariant subspace associated with the unstable eigenvalues; and (iii) exploit the constraints implied by initial conditions for predetermined variables, auxiliary initial conditions uncovered in computing the transition matrix, and the vectors of non-predetermined variables. This last step amounts to finding a system within which there is a matrix equivalent to H_{θ}, but is non-singular.

The task begins with the algorithm clustering all the zero rows associated with the H_{θ} block within H in the upper part of the matrix. Too many steps follow to be fully documented here. At previously noted, AIM uses the QR decomposition to find a transformation of the singular matrix H_{θ}:[3]

$$QR = H_{\theta} \tag{9}$$

3. There are other methods for solving systems of linear equations with singular lead matrices. Some, like Sims (1995) and Klein (1999) use the QZ, or generalized Schur decomposition. Binder and Pesaran (1995) advocate a martingale difference method and King and Watson (1995) document what they call a system reduction method. We have found the AIM algorithm to be very fast and reliable. Anderson (2000) points out that it is particularly amenable for use with large-scale models.

where Q is an orthogonal matrix and R is an upper-triangular matrix with the number of non-zero rows of R equalling the rank of H_θ. The algorithm then checks the rank of a suitably modified H_θ. If it is not of full rank, the matrix H is premultiplied by Q^T, an orthogonal transformation of equations (7) which annihilates as many zero rows of H_θ as possible. The process continues until H_θ is non-singular. When a non-singular lead matrix is found, its inverse is used to compute:

$$\Gamma = -H_\theta^{-1}[H_{-\tau}, \ldots, H_{\theta-1}] \tag{10}$$

which in turn appears at the bottom of the transition matrix $\tilde{H}$ used in equations (8) above:

$$\tilde{H} = \begin{bmatrix} 0 & I_{\langle n(\tau+\theta-1)\rangle} \\ \multicolumn{2}{c}{\Gamma} \end{bmatrix} \tag{11}$$

As shown here, the dimensionality of $\tilde{H}$ is shown as $n(\tau+\theta)$, the largest it can be; however, the annihilations mentioned above will usually reduce the size of $\tilde{H}$ resulting in some savings in computational time.

The remaining steps of the algorithm involve testing the eigenvalues of $\tilde{H}$ to ensure that they satisfy the Blanchard-Kahn (1980) conditions for the existence of a unique saddlepoint equilibrium. To speed the inversion of the sort of large, sparse matrices conventionally used in macroeconomic models, the algorithm makes use of sparse matrix code.

3. Remarks

There are three things worth noting about the above. First, notice that the computation of the solution does not require any hill climbing, nor does it require iteration over matrix Ricatti equations. Instead, the computation of costate equations does the work. Second, there is no need to identify and order the non-predetermined variables in advance and no need to write out the state-space representation. Third, our method of computing optimal control solutions, by exploiting the AIM algorithm, is ideally suited for use with large-scale models. These are advantages over other methods, including Söderlind (1999).

4. Two Examples

4.1 A small sticky-price model

As a first, small-scale example of the proposed method, we solved for the optimal rule for three sets of preferences for a standard sticky-price model of the business cycle, borrowed from Clarida, Gali and Gertler (1998). The model is very similar to most others in this field, including King and Wolman (1996), Woodford (1996) and McCallum and Nelson (1999). Prior to control, the model consists of two equations:

$$\pi_t = \delta E[\pi_{t+1} | \Omega_t] + \mu y_t \tag{12}$$

$$y_t = E[y_{t+1} | \Omega_t] - (1/\sigma)(r_t - E[\pi_{t+1} | \Omega_t]) \tag{13}$$

where y_t is the output gap, π_t is inflation, $1/\sigma$ is the elasticity of intertemporal substitution, r_t is the nominal federal funds rate;. See Clarida *et al.* (1998) for a detailed discussion of the model; we use the same parameterization as they do: $\delta = 0.99$, $\mu = 0.3$, $\sigma = 1$.

For present purposes, the most notable aspect of this model is that it is entirely forward looking; that is, it contains no predetermined elements whatsoever. This means that in the absence of an instrument smoothing motivation or a cross-product term in the objective function, the optimal rule is a function solely of costate variables. We solved for optimal rules with a mild penalty of 0.1 on the change in the federal funds rate, and with no discounting; i.e., $\beta = 1$, and three different relative weightings of output stabilization relative to inflation stabilization. Placing a small penalty on instrument variability prevents the coefficients of the optimal rule from becoming fantastically large and with them, the implied variability of the funds rate itself. The results are shown in Table 1 below.[4]

Because the solution method described above augments the original model with costate equations, the reduced form for the optimal rule--that is, the recalculation of the costates in terms of lags of state variables--can be easily done.

4. The weights on output gap variability and inflation variability are constrained to sum to unity without loss of generality. Solutions for preferences that do not include smoothing and substitutions of the costate variables as functions of lagged state variables are available from the authors on request.

4.2 The FRB/US model

Our second example is the optimal rule for a linearized version of FRB/US, the Board of Governors' large-scale rational expectations model of the U.S. economy.[5] With more than 300 equations, the optimal control rule for FRB/US ends up containing more than 350 arguments. Clearly this is a computational task of substance. It can be solved--from parsing the model code to the computation of all of the parameters--in 20 minutes on a 300Mhz UNIX SparcStation. While this is a significant amount of time, it has to be considered in context: Levin *et al.* (1999) did not compute optimal control rules with any of the models they compared, including FRB/US and smaller models, "due to [the] computational costs" associated with the larger models (p. 279). In fact, we have found the computation of optimal control rules with FRB/US by iterating on matrix Ricatti equations to be infeasible.

The computed parameters for a rule as complex as the optimal control rule for FRB/US are not easily interpreted and so we omit any discussion of them here. Instead, we compare the performance of optimal control rules for FRB/US with so-called "simple optimal rules"--rules that are optimal conditional on a restricted number of arguments to the rule. Tetlow and von zur Muehlen (1999), Levin *et. al.* (1999) and Williams (1999) have all argued, in one context or another, that adding parameters to a policy rule beyond the best three adds little to the performance of policy, on average.

Figure 1 below compares the optimal control rule with the performance of an extended Taylor rule:

$$R_t = \alpha_R R_{t-1} + (1-\alpha_R)(rr_t^* + \tilde{\pi}_t) + \alpha_y y_t + \alpha_\pi(\tilde{\pi}_t - \pi^*) \tag{14}$$

where R is the nominal federal funds rate, rr^* is the equilibrium real interest rate, y_t is the output gap, and $\tilde{\pi}_t$ is a four-quarter moving average of the inflation rate, and π^* is the target rate of inflation. Both rr^* and π^* can be set equal to zero without loss of generality. Equation (14) is one of the policy rules explored at length by Williams (1999).

It is important to note that policy rules like equation (14), which contain current dated arguments, are implicitly including many more than just three state variables. This is because nei-

5. See, Brayton and Tinsley (eds.)(1996), Brayton *et al.* (1997) and Reifschneider *et al.* (1999) for discussions on various aspects of the FRB/US model. Bomfim *et al.* (1997) and Williams (1999) explore some applications with the model.

ther y_t nor $\tilde{\pi}_t$ are state variables of the system; by including these variables in the rule, we are implicitly including all of the state variables that help predict these two variables, albeit in a constrained fashion.

The figure shows the two rules' performance in the form of two efficient policy frontiers; that is, curves along which it is impossible to find an alternative parameterization of the rule in question that can decrease the asymptotic standard deviation of inflation without increasing the standard deviation of the output gap.[6] The loss function used to compute the rules is the same one as in the small-scale example above; performance is measured using the variance-covariance matrix of model residuals computed over the period from 1966Q1 to 1998Q2.

The difference in performance between the two rules is fairly significant--probably more than one might have believed given the analysis Williams (1999), Levin *et al.* (1999) and Tetlow and von zur Muehlen (2000). For example, points A and A' on the figure show the performances of the rules for a 'strong' inflation targeter--a policymaker whose loss-function weights on inflation and output variability are 0.95 and 0.05 respectively. Constraining the policy rule to be of the form of equation (14) increases the asymptotic standard deviation of the output gap by about 0.4 percentage points and the standard deviation of inflation by 0.1 percentage points, relative to the optimal control solution. Given that the entire length of the optimal rule frontier is only a bit more than a percentage point in output variability and about half that in inflation variability, this should be seen as a noteworthy deterioration. At points B and B', the weights in the loss function on output and inflation variability are both 0.50; points C and C' represent the preferences of a 'weak' inflation targeter, one whose preference weightings are the opposite of the strong authority's.

Whatever the merits of the graphical representation of the cost of parsimony in Figure 1, it has the limitation that it covers only two dimensions, inflation and output, when the loss function contains a third element, the change in the federal funds rate. Figure 2 takes a broader look at the performance of the two rules by examining the computed loss of each as a function of the weight on output variability in the loss function (and implicitly of unity minus that weight on inflation variability). In particular the figure shows the *extra* loss of the 3-parameter rule relative to the loss of the optimal rule, measured in percent of the optimal loss. Thus, the height of the line shows how much more loss than is optimal must be borne by an authority that uses the 3-parameter rule.

---

6. Or the change in the federal funds rate which is also an argument to the loss function. We do not show the change in the funds rate here in order to simplify the figure. We return to this issue below.

Evidently, if the monetary authority is strong (in the sense described above), the cost of parsimony in the design of monetary policy rules is not particularly large; however, the more the authority's preferences lie toward stabilizing output, the more costly it is to rely on simple rules, even if those simple rules have been parameterized optimally. The large number and variety of channels through which shocks affect aggregate demand in FRB/US means that policy operating on a crude aggregation of these shocks--the output gap itself--is not very efficient for stabilizing output going ahead.[7] Controlling inflation, since it responds largely to *aggregate* demand, and not to the disaggregated components of demand, is more straightforward.

The observation that simple rules may do significantly worse than optimal rules suggests an interpretation for the reluctance of central banks to announce a particular rule: a central bank that cares about real variability needs to decompose output into the channels and shocks that determine output variability going ahead. Committing to a simple rule may not be adequate for this task.

7. This interpretation is buttressed by the fact that for the model of Tetlow and von zur Muehlen (2000), which uses a simple aggregate IS equation for the determination of output, the curve in Figure 2 slopes downward rather than upward.

5. References

Anderson, G.S., 2000, A reliable and computationally efficient algorithm for imposing the saddle-point property in dynamic models, unpublished manuscript, Division of Monetary Studies, Board of Governors of the Federal Reserve System.

Anderson, G.S., Moore, G, 1985, A linear algebraic procedure for solving linear perfect foresight models, Economics Letters, 17: 247-252.

Backus, D., Driffill, J., 1986, The consistency of optimal policy in stochastic rational expectations models, Discussion paper no. 124, CEPR, London.

Binder, M., Pesaran, M.H., 1995, Multivariate rational expectations models and macroeconometric modeling: a review and some new results, In: Handbook of Applied Econometrics, Oxford, Blackwell.

Blanchard, O.J., Kahn, C.M., 1980, The solution of linear difference equations under rational expectations Econometrica 48, 1305-1311.

Bomfim, A., Tetlow, R., von zur Muehlen, P., Williams, J. 1997, Expectations, learning and the costs of disinflation: experiments using the FRB/US model, Finance and Economics Discussion Series no. 1997-42, Board of Governors of the Federal Reserve System.

Brayton, F., Tinsley. P., eds., 1996, A guide to FRB/US: a macroeconometric model of the United States, Finance and Economics Discussion Series no. 1996-42, Board of Governors of the Federal Reserve System.

Brayton, F., Mauskopf, E, Reifschneider, D., Williams, J.C. 1997, The role of expectations in the FRB/US macroeconomic model, Federal Reserve Bulletin, 83, 227-245.

Chow, Gregory C., 1997, Dynamic Economics: Optimization by the Lagrange Method, Oxford U. Press, New York.

Clarida, R, Gali, J., Gertler, M., 1998, Monetary policy rules and macroeconomic stability: evidence and some theory, working paper no. 6442, NBER, Cambridge, MA.

King, R.G., Watson, M.W, 1995, System reduction and solution algorithms for singular linear difference systems under rational expectations, unpublished manuscript,

King, R.G., Wolman, A.L., 1996, Inflation targeting in a St. Louis model of the 21st century, working paper no. 5507, NBER, Cambridge, MA.

Klein, P., 1999, Using the generalized schur form to solve a multivariate linear rational expectations model. Journal of Economic Dynamics and Control, forthcoming.

Levin, A, Wieland, V, Williams, J.C., 1999, Robustness of simple monetary policy rules under model uncertainty. In: Taylor, J.B. (ed.) Monetary Policy Rules, North-Holland, Amsterdam.

Levine, P, Currie, D., 1987, The design of linear feedback rules in linear stochastic rational expectations models, Journal of Economic Dynamics and Control, 11, 1-28.

McCallum, B.T, Nelson E., 1999, Performance of operational policy rules in an estimated semi-classical structural model, In: Taylor, J.B. (ed.) Monetary Policy Rules, North-Holland, Amsterdam.

Reifschneider, D., Tetlow R, Williams, J.C., 1999, Aggregate disturbances, monetary policy and the macroeconomy: the FRB/US perspective, Federal Reserve Bulletin, 25, 1-19.

Rudebusch, G. D., Svensson, L.E.O., 1999, Policy rules for inflation targeting, in: Taylor, J.B. (ed.) Monetary Policy Rules, North-Holland, Amsterdam.

Sims, C.A., 1995, Solving linear rational expectations models, unpublished manuscript, Yale University.

Söderlind, P., 1999, Solution and estimation of RE macromodels with optimal policy, European Economic Review 43, 813-823.

Tetlow, R.J., von zur Muehlen, P., 2000, Simplicity versus optimality: the choice of policy rules when agents must learn, Journal of Economic Dynamics and Control, forthcoming.

Williams, J. C., 1999, Simple rules for monetary policy, Finance and Economics Discussion Series paper no. 1999-12 (February).

Woodford, M., 1996, Control of the public debt: a requirement for price stability? working paper no. 5684, NBER, Cambridge, MA.

Table 1
Optimal Policy Rules for the Clarida, Gali & Gertler Model

rule argument	Q_y **: loss function weight on** y		
	0.15	**0.50**	**0.85**
r_{t-1}	**0.335**	**0.238**	**0.209**
$\lambda_{y,\,t-1}$	**2.257**	**0.934**	**0.269**
$\lambda_{\pi,\,t-1}$	**0.102**	**-0.178**	**-0.464**

The loss function is: $E\sum_{i=0}^{\infty}[y'Q_y y + \pi'(1-Q_y)\pi + \Delta r'0.1\Delta r]$ with weight $Q_y \in \{0.15, 0.50, 0.85\}$. The corresponding optimal policy rules are of the form: $r_t = a_1 r_{t-1} + a_2\lambda_{y,\,t-1} + a_2\lambda_{\pi,\,t-1}$. Without smoothing, the term in $a_1 r_{t-1}$ drops out.

Figure 1

Optimal and Constrained Optimal Policy Frontiers for FRB/US

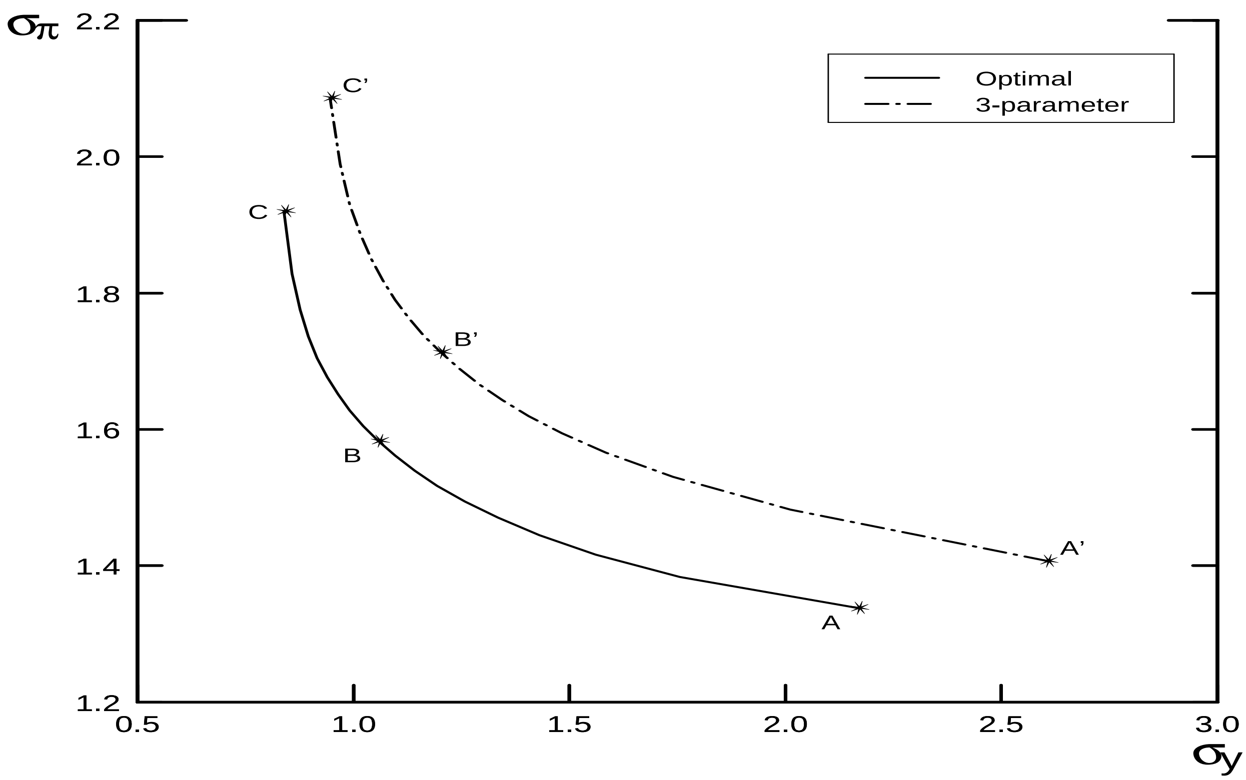
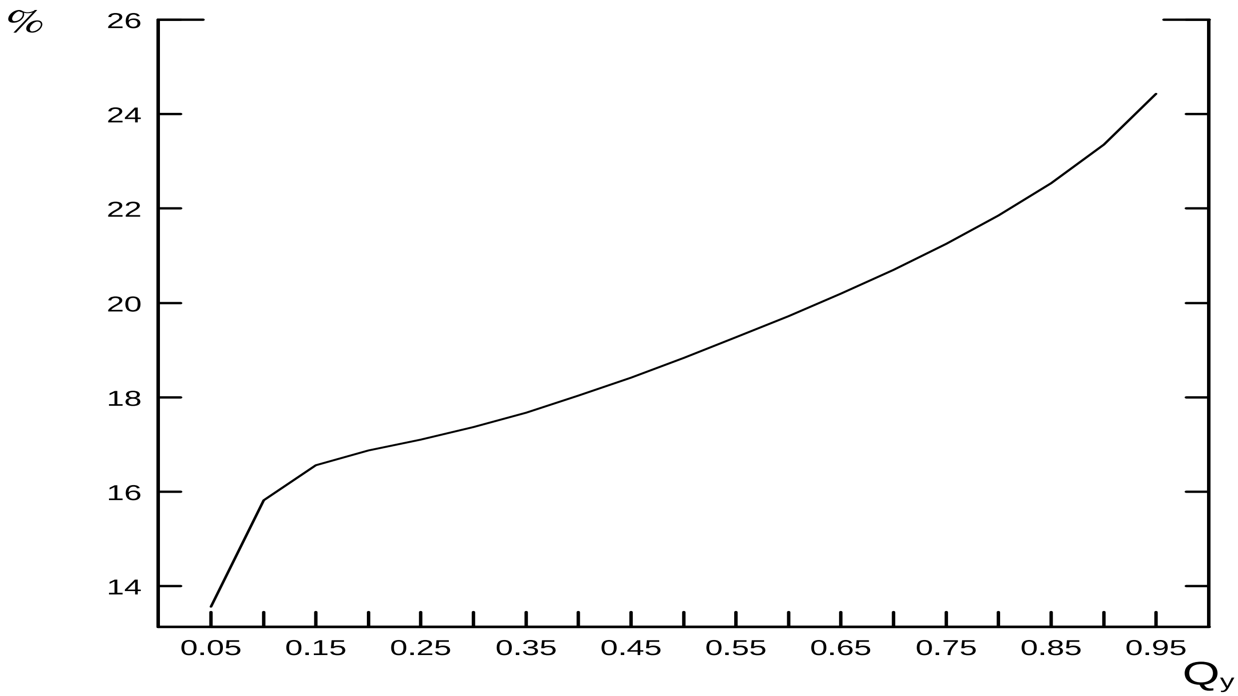


Figure 2

The Cost of Parsimony
3-parameter rule relative to the optimal rule
(percent of optimal rule loss)



$$J = Q_y\sigma_y^2 + (1-Q_y)\sigma_\pi^2 + .1\sigma_{\Delta r}^2$$