A Simple Permanent Income Model: Comparing Numerical Methods

by Christopher A. Sims and the 1998 Econ. 526b class

I. The Model

A. Standard Form

$$\max_{\{C_t, W_t\}_0^\infty} E\left[\sum_{t=0}^{\infty} \beta^t \cdot \left(C_t - \tfrac{1}{2} C_t^2\right)\right] \tag{1}$$

subject to

$$W_t = R \cdot \left(W_{t-1} - C_{t-1}\right) + Y_t \tag{2}$$

$$\lim_{t\to\infty} E\beta^{t/2} W_t = 0 \tag{3}$$

or equivalently

$$A_t = RA_{t-1} - C_t + Y_t \tag{4}$$

$$\lim_{t\to\infty} E\beta^{t/2} A_t = 0 \ . \tag{5}$$

As is well known, this leads in the special case $R\beta = 1$ to the first order condition

$$C_t = E_t C_{t+1} \tag{6}$$

and, in the further specialized case of i.i.d. Y_t, to the decision rule

$$C_t = \left(1-\beta\right) \cdot W_t + \beta \overline{Y} \ . \tag{7}$$

Explicit solution for W in terms of Y alone gives us

$$W_t = W_{t-1} + Y_t - \overline{Y} \ . \tag{8}$$

The rule of thumb approximation to the implications of permanent income consumption theory: "Consumption and wealth are random walks," that is, they have zero mean i.i.d. increments.

B. Our modified model

Replace (3) with

$$C_t \le W_t \tag{9}$$

or (5) with

$$A_t \ge 0 \ . \tag{10}$$

These constraints have better economic motivation. The model is no longer LQ, because the constraints can be expected to sometimes bind, sometimes not. On the other hand, it has a concave, continuous objective function and a convex constraint set, so we know it has a solution, and that a solution to the FOC's will be an optimum. The FOC changes to the form

$$C_t = E_t C_{t+1} - \mu_t, \tag{11}$$

where μ_t is the random Lagrange multiplier on (9), which is non-negative and zero when (9) does not bind.

It is immediately clear that for extreme values of W this model has solutions quite different from the standard model. The standard decision rule (7) and the constraint (9) conflict for low values of W, in particular for values of W less than $\overline{Y}$. Furthermore, it is easy to see that, since the objective function has a satiation point in its one-period utility function, it can never be optimal to choose $C_t \geq 1$. Doing so can only decrease current utility, and its effect on the opportunity set for future C's can only be to expand it: The increased savings this period makes more wealth available for future periods, which can never be harmful. Thus the standard decision rule will certainly conflict with optimality when W is very large, in particular when $W > (1-\beta\overline{Y})/(1-\beta)$.

A little further reflection tells us that the time path of C differs qualitatively from that in the standard case. Equation (11) tells us that C_t is a sub-martingale, because of the non-negativity of μ. The fact that C can never exceed one tells us further that it is a sub-martingale bounded above. The martingale convergence theorem then tells us that it converges almost surely. If C converges to any constant $\overline{C}$, W eventually starts to behave to an arbitrarily good approximation like a solution to the unstable stochastic difference equation $W_t = RW_{t-1} - R\overline{C} + Y_t$. So long as $R > 1$ and Y has a non-degenerate distribution, W therefore eventually grows arbitrarily large or becomes less than $\overline{C}$. The latter violates a constraint, and is thus not possible. If W grows arbitrarily large, eventually it must be possible to improve utility without risk of future consequences by "eating" part of W, unless $\overline{C}$ is one, in which case increasing C does not increase utility[1]. Thus the optimal path of C converges, as an ordinary sequence of real numbers, to the satiation level of 1. [And contrary to the suggestions from the floor during the presentation of this model 4/2, which I mistakenly acceded to, the result does not depend on any restriction on initial wealth or on the support of the distribution of Y.] This means that the inequality (9) binds only finitely many times, so that eventually μ is zero and stays there. The FOC then says that in

---

[1] If Y has a deterministic lower bound, say Y_L (even if this is negative), the argument is easy. If $(R-1)W$ exceeds $1-Y_L$, then the interest on wealth alone can finance satiation consumption forever, even if Y is at its lower bound every period, so once W gets this large, it cannot be optimal to consume less than 1. If Y does not have a lower bound, the argument is more technical, but is the same as that required to justify the necessity of the standard transversality condition, $E\left[\beta^t(1-C_t)W_t\right] \underset{t\to\infty}{\to} 0$, which is violated in any equilibrium in which C has any probability of converging to a constant less than one.

the long run the standard form of the FOC, implying C is a martingale, is correct, but C is a convergent martingale, not a random walk.

II. Solutions

There is no analytic solution for the decision rule in the non-standard version of the problem. It is interesting to ask the question, "In what sense and over what range is the simple solution to the standard problem a good approximation to the truly optimal solution?" People often think of a quadratic objective function as a good local approximation to an arbitrary smooth objective function and of the standard solution as therefore a good local approximation to behavior in an arbitrary smooth model. Can we justify these ideas?

A. Numerical Methodology Disputes

In our problem, we will iteratively improve an estimate of the optimal consumption rule $\gamma(W)$, measuring accuracy by whether the constraints are satisfied and whether (11) is satisfied. Of course (11) involves an expectation, so checking it is not straightforward. After substitution of γ for C, (11) becomes

$$\gamma(W) = E\left[\gamma\left(R\left(W-\gamma(W)\right)+Y\right)\right]+\mu(W). \tag{12}$$

Note that the presence of μ means that in effect the FOC is satisfied trivially when $\gamma(W)=W$. Also, note that the only random element in (12) is Y, so that it is this random variable with respect to which we take the expectation.

1. Judd-Marcet/Marshall-Rust

a) Discretization vs. parametric approximation

The advantage of parametric methods is that they allow us to exploit knowledge that solutions are smooth in order to conserve on the dimension of our search. In this model, as in most macro models, we do not expect discontinuities, or even lack of differentiability, in the decision rule. To approximate it with increasingly refined discrete approximations, then, tends to produce increases in accuracy incommensurate with the increased computational effort, since at some point the new solutions start looking simply like interpolations of the old ones. On the other hand, excessively smooth parametric approximations can produce large approximation error, perhaps concealing it in reasonable-looking shapes of the solutions.

b) Monte Carlo integration vs. deterministic numerical integration

Given a γ and a W, evaluating the expectation in (12) is just a one-dimensional integration. Furthermore, with a polynomial or spline parameterization of γ and a standard form for the distribution of Y, the integration can be done analytically. However, the code to do the analytic integration is moderately complex if a spline approximation is used or if the exact $C \le W$ constraint and/or $C \le 1$ boundary condition are imposed by truncating γ rather than by building them in the parameterization of γ or by treating deviations from them as part of what is penalized in the accuracy criterion. Deterministic numerical integration of smooth one-dimensional functions is fast and accurate, and there are standard programs for it. Monte-Carlo integration is

scarcely any easier for a problem this simple. It does insulate us against the possibility that some aspect of the function we choose to integrate may interact with a systematic algorithm for numerical integration to produce unusually inaccurate results. Of course, a Monte Carlo draw can always, randomly, give us bad results also. But we can characterize the probabilities of those bad results. In principle we could do the same for deterministic numerical integration, but only by an impracticably complicated, analysis of our subjective beliefs about the class of functions we are integrating and about their interaction with the numerical integration scheme. We choose Monte Carlo integration here because it makes the connection between Marcet/Marshall and Galerkin methods more transparent.

c) Parameterize Euler equation expectations or decision rules

Though Marcet and Marshall's approach is usually called "parameterized expectations", this is not a central distinction between it and other approaches. In this problem, for example, parameterized expectations suggests replacing the right-hand-side of (11) with a parameterized function of the state, W. But this is just what we do when we postulate $C = \gamma(W)$.

d) Check numerical errors in simulation path space or in decision rule space

This is a major distinction. Any method will have to check the validity of (12), but this equation is supposed to hold at every W. Since W takes on a continuum of values we cannot check the equation at every W. The straightforward approach is to form a grid of W values, check (12) at each point on the grid, and somehow aggregate the errors to form an accuracy criterion. What Marcet and Marshall suggest instead is that we choose only an initial value of W, then use our random draws of Y not to generate a sample of values of the right-hand-side of (12) at each W on a grid, but instead use them to generate a long sequence of W's, recursively. That is, they suggest we simulate a time path of the model and collect the discrepancies between the right-hand-side of (12) (without the E operator) and the left-hand-side.

Both approaches leave us with the problem of aggregating all the individual errors into an error criterion. Marcet's and Marshall's initial suggestion was to treat the R^2 of a regression of the simulated Euler equation residuals at $t+1$ on functions of the state at t as the error criterion. Judd's was to form a k-dimensional error criterion, where k is the dimension of the parameterization of γ, by forming the cross-products of mean errors at each point on the W grid with functions (e.g. powers) of elements of the W grid itself.

The Marcet/Marshall approach has the advantage that it automatically focuses attention on accuracy of (12) at a set of W's that are actually encountered frequently in a simulation. The Judd approach requires that we pick a W grid without knowing what the model's solution is or what its simulated path looks like. It can therefore waste effort on regions of W-space that are almost irrelevant in simulation, or make the grid too sparse in regions where nearly all W values concentrate in simulations. On the other hand, some uses of a solved model may require having accuracy in γ at W values that are rare in a simulation. For such uses, the Marcet-Marshall approach may be inadequate.

Each approach can be adjusted, though. Simulations can be repeatedly restarted from initial values in a range of interest. A fixed W grid can be adjusted and the solution retuned if initial simulations suggest that the grid is too coarse in some area.

In the results below, only computations based on a fixed W grid are shown.

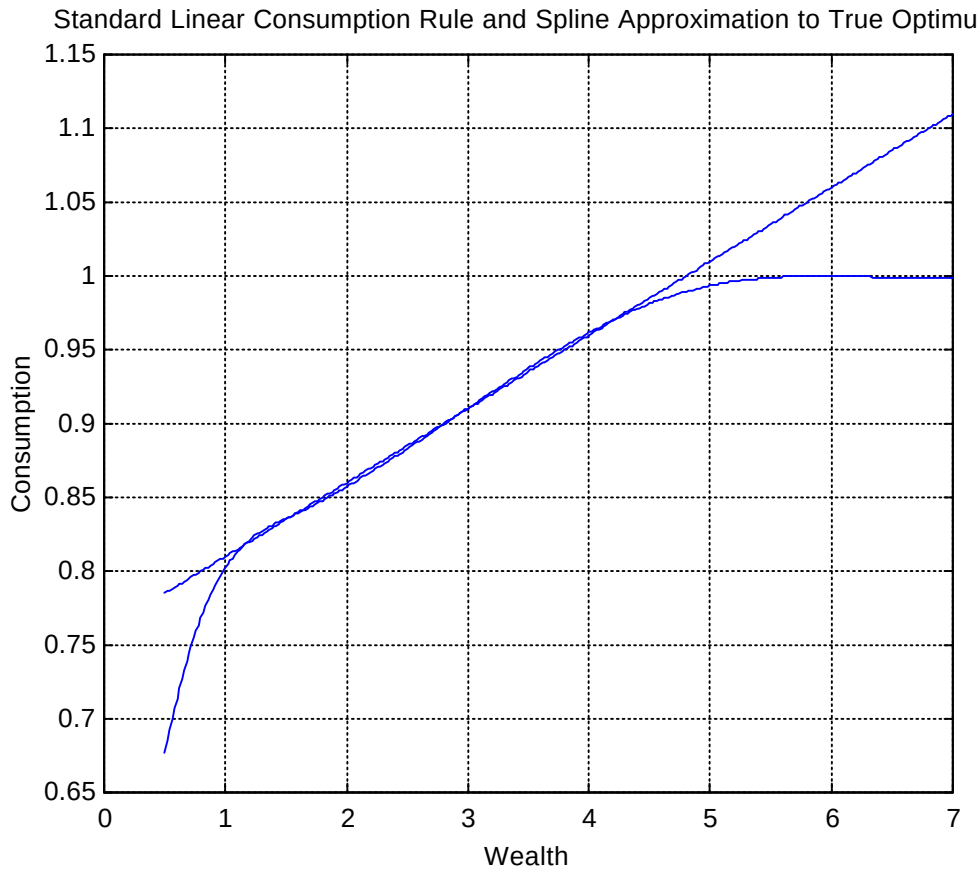
e) Fixed-point vs. gradient-based optimization

Marcet-Marshall originally suggested a fixed point method. In particular, they suggested that at each iteration one should estimate the parameters of γ by a nonlinear least squares regression of the simulated values of the right-hand-side of (12) on γ. At the true solution, this should give parameter values that match those of the γ that generated the simulated data, if the simulation is long enough. One could reduce the number of iterations required by instead using a standard gradient-based method to update the parameters of γ, treating the R^2 as a criterion to be minimized. As Judd points out, experience seems to suggest that it generally works better to formulate the task as an equation-solving problem rather than as a minimization. Thus one could, while staying with Marcet and Marshall's simulation approach, solve for the parameter vector that set to zero the sample cross products of the simulated Euler equation residuals with functions of the simulated W's.

In the results below, all equation-solving was done with a gradient-based method.

Thus Judd is *RRRRR*, Marcet/Marshall *RLLLL*. Rust, though more eclectic, may usually be *LLRRL*. Most of the RBC literature represents *LRRRL*. Judd has emphasized that he recommends "Galerkin methods", which underlines (a) and (c), while Marcet and Marshall call their approach "parameterized expectations", which also emphasizes (a) and (c). However, there is a tendency for people to follow the examples of these computational leaders whole hog, in all dimensions (a)-(e) at once, when in fact these represent independent dimensions of choice in numerical methods. Furthermore, there are other dimensions of choice in method that are varied by each leader and hence are not controversial in the same way, yet which are perhaps just as crucial as these 5.

We will consider solutions to this surprisingly challenging simple model that use *RL_RR* methods. The blank space corresponds to the fact that (c) is empty in this model – there is no difference between parameterizing the decision rule and parameterizing the Euler equation expectations. Though our approach is basically parametric (R rather than L on (a)), the solution presented will be based on a spline parameterization, which, like discretized solutions, avoids the strong dependence between the behavior of the approximated function at distant points in the state space. In particular, we use a cubic spline with 4 knots, constraining the derivative to the left of the first knot to be one (as it represents the $C=W$ region), the level of the function at the last knot to be one, and the derivative at the last knot to be zero (reflecting the approximate $C=1$ behavior for large W). We pick a grid of W values running from .5 to 7 with increments of .1, and evaluate (12) for 200 $N(.8,.01)$ draws of Y at each W. We solve for spline knot points (6 free parameters, 4 W values and 2 unconstrained corresponding C values) that make the cross products of the mean discrepancy between left and right hand sides of (12) with the 0'th through 5^{th} powers of W zero. We choose $\beta = .95$. The result is displayed in the figure below.



Note that the standard model solution lies almost exactly on top of our more realistic model for wealths between about 1.2 and 4.4. The level of wealth at which the wealth constraint would have to override the standard decision rule is .8, which lies 4 standard deviations of Y below 1.2. Thus noticeable deviations from the standard rule appear at wealth levels high enough that there is still a negligible chance of the wealth constraint binding in the next period. The level of wealth at which earnings on wealth plus expected Y are enough to cover satiation consumption is 4.8. Thus noticeable deviations in the direction of increased savings at high wealth levels show up at about 4 standard deviations below the level (4.8) that would suffice to finance satiation forever under certainty. Consumption remains noticeably (at this scale) below satiation up to a wealth of about 5.5, seven standard deviations above the 4.8 level that, under certainty, would push C to 1.