

# CALCULATING AND USING SECOND ORDER ACCURATE SOLUTIONS OF DISCRETE TIME DYNAMIC EQUILIBRIUM MODELS

JINILL KIM, SUNGHYUN KIM, ERNST SCHAUMBURG, AND CHRISTOPHER A. SIMS

ABSTRACT. We describe an algorithm for calculating second order approximations to the solutions to nonlinear stochastic rational expectations models. The paper also explains methods for using such an approximate solution to generate forecasts, simulated time paths for the model, and evaluations of expected welfare differences across different versions of a model. The paper gives conditions for local validity of the approximation that allow for disturbance distributions with unbounded support and allow for non-stationarity of the solution process.

## 1. INTRODUCTION

It is now widely understood how to obtain first-order accurate approximations to the solution to a dynamic, stochastic general equilibrium model (DSGE model). Such solutions are fairly easy to construct and useful for a wide variety of purposes. They are likely to be accurate enough to be a basis for fitting the models to data, for example.

However, for some purposes first-order accuracy is not enough. This is true in particular for comparing welfare across policies that do not have first-order effects on the model's deterministic steady state, for example. It is also true for attempts to study asset pricing in the context of DSGE models. It is possible to assume directly that nonlinearities are

---

*Date:* August 3, 2003.

Discussions, and in some cases exchange of code testing results, with Fabrice Collard, Kenneth L. Judd, Robert Kollmann, Stephanie Schmitt-Grohé, and Martin Uribe have been useful to us. Kollmann has contributed to the Matlab code that implements the paper's algorithm. ©2003 by Jinill Kim, Sunghyun Kim, Ernst Schamburg and Christopher Sims. This material may be reproduced for educational and research purposes so long as the copies are not sold, even to recover costs, the document is not altered, and this copyright notice is included in the copies.

themselves small in certain dimensions as a justification for use of first-order approximations in these contexts; Woodford (2002) is an example of making the necessary auxiliary assumptions explicit. But the usual reliance on local approximation being generally locally accurate does not apply to these contexts.

It is therefore of some interest to have an algorithm available that will produce second-order accurate approximations to the solutions to DSGE's from a straightforward second-order expansion of the model's equilibrium equations, and this is an active area of recent research.

Kenneth Judd pioneered this field by using perturbation methods in solving various types of economic models<sup>1</sup>. Jin and Judd (2002) describe how to compute approximations of arbitrary order in discrete-time rational expectations models. They aim at providing a complete set of regularity conditions justifying the local approximations, and they discuss methods for checking the validity of the approximations. Others also have studied perturbation methods of higher than first order including Collard and Juillard (2000), Anderson and Levin (2002), and Schmitt-Grohé and Uribe (2002).

Kim and Kim (2003a) and Sutherland (2002) have developed a bias correction method that produce the same results as the second order perturbation method for certain welfare calculations, while requiring less computational effort than a full perturbation solution.

Several papers have applied the second-order perturbation method to dynamic general equilibrium models. Kim and Kim (2003b) used the second-order solution method to analyze welfare effects of tax policies in a two-country framework. In particular, they calculate the optimal degree of response for various tax rates to TFP shocks faced by each country. Welfare gains of tax policies are measured by conditional welfare changes from the benchmark case. Kollmann (2002) has analyzed the welfare effects of monetary policies in open economies using the software that has been developed along with this paper, and Bergin and Tchakarov (2002) have used it to examine the welfare effects of exchange rate risk.

---

<sup>1</sup>Judd (1998). For continuous time models see Gaspar and Judd (1997) as well.

This paper describes the algorithm for computing a second order approximation and shows how to apply it to calculating forecasts and impulse responses in dynamic models and to evaluating welfare in DSGE models. It points out some necessary regularity conditions for application of the method and discusses the sense in which the approximate solutions are locally accurate.

While much of the paper parallels others in this rapidly growing literature, this paper makes some new contributions. The rest of the literature in most cases begins from a formulation of the problem in which a partition of variables in the model into “states” and “controls” or “co-states” is assumed known. While in smaller models such a partition is often obvious, in larger models it can be unclear how to partition the variables into states and controls. The Matlab program `gensys.m`, implementing the approach described in Sims (2001), accepts model specifications that do not partition the variable list into predetermined and non-predetermined variables; instead it partitions disturbances into predetermined and non-predetermined categories. This approach is more natural in systems derived from equilibrium models, in which equation disturbances often fall neatly into these categories. In such models translating the list of predetermined disturbances into a corresponding list of predetermined variables (or, where necessary, new predetermined variables that are linear combinations of the original model’s variables) may not be easy. This paper extends that approach to second-order approximations.<sup>2</sup>

The “state-free” approach of `gensys.m` has the disadvantage that its output, while completely characterizing the dynamics in terms of the original variables, includes only its own artificial decomposition into states and co-states, which may be opaque. For some purposes it is important to have an intuitively appealing decomposition into states and co-states. We discuss how to do this, with the aid of another program, `gstate.m`, that uses the output of `gensys.m` or `gensys2.m` to test proposed state vectors and to provide guidance as to what a valid state vector must look like.

---

<sup>2</sup>King and Watson (1998) and Klein (2000) describe solution algorithms that handle the essentially the same class of models as Sims (2001), but presume that the list of predetermined variables is given.

Where the sense in which accuracy of local expansions is claimed has been made explicit in the literature, it has for the most part (Jin and Judd, 2002, most prominently) focused on accuracy of the function mapping state variables to co-states. It has also tended to assert as regularity conditions almost-sure boundedness of stochastic disturbances and stationarity of the dynamic model being studied. These assumptions allow strong claims to be made about approximation accuracy, but they are disquieting for most DSGE modeling applications. Models with unit roots, or even mild explosiveness, are not uncommon in macroeconomics, and models with near-unit roots are the rule. Often disturbance distributions with unbounded support seem more realistic than any particular truncation to bounded support. If perturbation methods break down, or are at the edge of their domain of applicability, for such models, they might seem to be unattractive for many of the models to which they have in fact been applied.

In this paper we argue that boundedness of shocks and stationarity of the model are not essential to the validity of perturbation methods. For their main applications so far, perturbation methods can be shown to produce results that are in a natural sense locally accurate, without the invocation of the dubious stationarity and boundedness assumptions.

There is little explicit discussion in the literature of how to use higher order perturbation approximations in constructing simulations, forecasts, and welfare evaluations. We show that some apparently obvious approaches to these tasks in fact result in an accumulation of “garbage” high-order terms that can make accuracy deteriorate. We lay out an algorithm that always produces stationary second-order accurate dynamics whenever the first-order dynamics are stable.

The Matlab code that was built along with this paper is available at <http://eco-072399b.princeton.edu/yftp/gensys2/>, where the current version of this paper will also be found.

## 2. THE GENERAL FORM OF THE MODEL

We suppose a model that takes the form

$$(1) \quad K \begin{pmatrix} w_t & w_{t-1} & \sigma \varepsilon_t \\ n \times 1 & n \times 1 & m \times 1 \end{pmatrix} + \Pi \sigma \eta_t = 0, \quad p \times 1$$

where  $E_t \eta_{t+1} = 0$  and  $E_t \varepsilon_{t+1} = 0$ .<sup>3</sup> The equations hold for  $t = 0, \dots, \infty$ , as does the  $E_t \varepsilon_{t+1} = 0$  condition. The disturbances  $\varepsilon_t$  are exogenously given, while  $\eta_t$  is determined as a function of  $\varepsilon$  when the model is solved, if the solution exists and is unique. Note that because there is no assumption at all about  $\eta_0$ , it is a free vector that is likely to make certain linear combinations of the equations tautological at the initial date.

The scale factor  $\sigma$  is introduced to allow us to shrink the distribution of  $\varepsilon_t$  toward zero as we seek a domain of validity for our local approximation. The distribution of  $\varepsilon_t$  itself is assumed to be constant across time  $t$  and invariant to changes in  $\sigma$ , so that in particular it has a fixed covariance matrix  $\Omega$ .

The equation system could be written equivalently as

$$(2) \quad Q_1 K(w_t, w_{t-1}, \sigma \varepsilon_t) = 0$$

$$(3) \quad E_t [Q_2 K(w_{t+1}, w_t, \sigma \varepsilon_{t+1})] = 0,$$

where  $Q_1$  is any matrix such that  $Q_1 \Pi = 0$  and  $[Q_1', Q_2']$  is a full rank square matrix. The “forward-shift” of the expectational block reflects the absence of any restriction on  $\eta_0$ .

We assume that the solution will imply that  $w_t$  remains always on a stable manifold, defined by  $H(w_t, \sigma) = 0$  and satisfying

$$(4) \quad \left\{ \begin{array}{l} H(w_t, \sigma) = 0, H(w_{t+1}, \sigma) = 0 \text{ a.s.} \\ n_u \times 1 \quad n \times 1 \end{array} \right. \quad \text{and} \quad Q_1 K(w_{t+1}, w_t, \sigma \varepsilon_{t+1}) = 0 \text{ a.s.} \} \\ \Rightarrow E_t [Q_2 K(w_{t+1}, w_t, \sigma \varepsilon_{t+1})] = 0.$$

---

<sup>3</sup>This form is more general than it might seem. See Sims (2001) for examples showing how models with explicit expectations operators, including lagged expectations, can be cast into this form.

We consider expansion of the system about a deterministic steady state  $\bar{w}$ , i.e. a point satisfying  $K(\bar{w}, \bar{w}, 0) = 0$ . We do not need to assume the steady state is unique, so the situation arising in unit root models, where there is a continuum of steady states, is not ruled out.

We also assume that the nonlinear system (1) is formulated in such a way that its first-order expansion characterizes the first-order behavior of the deterministic solution. That is, we assume that solving the first-order expansion of (1) about  $\bar{w}$ ,

$$(5) \quad K_1 dw_t = -K_2 dw_{t-1} - K_3 \sigma \varepsilon_t + \Pi \eta_t,$$

as a linear system results in a unique stable saddle path in the neighborhood of the deterministic steady state. If so, this saddle path characterizes the first-order behavior of the system. We assume further that  $H_1(\bar{w}, 0)$  is of full row rank, so that the first-order character of the saddle path is determined by the first-order expansion of  $H$ .<sup>4</sup>

The system (1) has the second-order Taylor expansion about  $\bar{w}$

$$(6) \quad K_{1ij} dw_{jt} = -K_{2ij} dw_{j,t-1} - K_{3ij} \sigma \varepsilon_{jt} + \Pi_{ij} \eta_{jt} \\ - \frac{1}{2} (K_{11ijk} dw_{jt} dw_{kt} + 2K_{12ijk} dw_{jt} dw_{k,t-1} + 2K_{13ijk} dw_{jt} \sigma \varepsilon_{kt} \\ + K_{22ijk} dw_{j,t-1} dw_{k,t-1} + 2K_{23ijk} dw_{j,t-1} \sigma \varepsilon_{kt} + K_{33ijk} \sigma^2 \varepsilon_{jt} \varepsilon_{kt}),$$

where we have resorted to tensor notation. That is, we are using the notation that

$$(7) \quad A_{ijk} B_{mnjq} = C_{ikmnq} \quad \Leftrightarrow \quad c_{ikmnq} = \sum_j a_{ijk} b_{mnjq}.$$

where  $a, b, c$  in this expression refer to individual elements of multidimensional arrays, while  $A, B, C$  refer to the arrays themselves. As special case, for example, ordinary matrix multiplication is  $AB = A_{ij} B_{jk}$  and the usual matrix expression  $A'BA$  becomes  $A_{ji} B_{jk} A_{km}$ .

---

<sup>4</sup>This assumption on  $H$  is not restrictive so long as there is a continuous, differentiable saddle manifold. However there are models — some asset pricing models, for example — in which the first order approximation does not deliver determinacy, but higher-order terms do. The algorithms suggested here cannot handle models of this type.

Note that we are distinguishing the array  $K_{mij}$  of first derivatives from the array  $K_{mnik}$  of second derivatives only by the number of indexing subscripts the two arrays have.

### 3. REGULARITY CONDITIONS

Because we are taking first and second derivatives and because we are expanding about the steady state  $\bar{w}$ , it is clear that we require existence of first and second derivatives of  $K$  at  $\bar{w}$ . We have also directly assumed that the first order behavior of  $K$  near  $\bar{w}$  determines  $H(\cdot, 0)$ . In order to make our local expansion in  $dw$ ,  $\sigma\varepsilon$ , and  $\varepsilon$  work, we will need that  $H(w, \sigma)$  is continuous and twice-differentiable in both its arguments.

It may seem that these are all standard assumptions on the degree of differentiability of the system near  $\bar{w}$ . Consider what emerges, though, when we split the system into expectational and non-expectational components as in (2)-(3). If we replace (3) with its second-order expansion and take some expectations explicitly, we arrive at

$$(8) \quad E_t \left[ Q_2 (K_{1ij} dw_{j,t+1} + K_{2ij} dw_{jt} + \frac{1}{2} (K_{11ijk} dw_{j,t+1} dw_{k,t+1} + 2K_{12ijk} dw_{j,t+1} dw_{k,t} + K_{22ijk} dw_{j,t} dw_{k,t} + K_{33ijk} \Omega_{jk} \sigma^2)) \right] = 0,$$

and find ourselves needing to assert that  $\varepsilon_t$  has finite second moments, which is not a local property. That is, if  $\varepsilon_t$  does not have second moments, shrinking  $\sigma$  will not make  $\sigma\varepsilon_t$  have finite second moments. The same point applies to (3) in its original nonlinear form. If it is to be differentiable in  $w_t$  and  $\sigma$ , we will in general need to impose restrictions on the distribution of  $\varepsilon_t$ . Jin and Judd (2002) have an example of a model in which some apparently natural choices of a distribution for  $\varepsilon_t$  imply that  $E_t [Q_2 K(w_{t+1}, w_t, \sigma\varepsilon_{t+1})]$  is discontinuous in  $\sigma$  at  $\sigma = 0$ , even though  $K$  has plenty of derivatives at the steady state.

### 4. SOLUTION METHOD

The solution we are looking for can be written in the form

$$(9) \quad w_t = F^*(w_{t-1}, \sigma\varepsilon_t, \sigma).$$

Because we know the saddle manifold characterized by  $H$  exists and that  $H_1(\bar{w}, \sigma)$  has full row rank  $n_u$ , we can use  $H$  to express  $n_u$  linear combinations of  $w$ 's in terms of the remaining  $n_s = n - n_u$ . Let the  $n_s$  linear combinations of  $w$ 's chosen as “explanatory” variables in this relation be

$$(10) \quad y_t = \underset{n_s \times n}{\Phi} w_t.$$

Then the solution (9) can be expressed equivalently, in a neighborhood of  $\bar{w}$ , as

$$(11) \quad y_t = \Phi F^*(w_{t-1}, \sigma \varepsilon_t, \sigma) = F(y_{t-1}, x_{t-1}, \sigma \varepsilon_t, \sigma)$$

$$(12) \quad \underset{n_u \times 1}{x_t} = h(y_t, \sigma),$$

where (12) is just the solved version of the  $H = 0$  equation that characterizes the stable manifold. Here of course  $x$ , like  $y$ , is a linear combination of  $w$ 's.

The appearance of  $x_{t-1}$  in (11) may seem redundant, since along the solution path we will have  $x_t = h(y_t, \sigma)$ , but at the initial date the lagged  $w$  vector might not satisfy this restriction. This is likely in a growth model with multiple types of capital, for example, where there may be optimal proportions of capital of different types, but no physical requirement that the initial endowments are in these proportions.<sup>5</sup>

The solution method for linear rational expectations systems described in Sims (2001) begins by applying linear transformations to the list of variables and to the equation system to produce an upper triangular block recursive system. In the transformed system, the unstable roots of the system are all associated with the lower right block,  $\eta_t$  does not appear in the upper set of equations in the system,<sup>6</sup> and the upper part of the equation system is normalized to have the identity as the coefficient matrix on current values of the upper part of the transformed variable matrix. In other words, by applying to the equation system the

---

<sup>5</sup>See section 5 below for further discussion of this point.

<sup>6</sup>It may not be possible in fact to eliminate  $\eta_t$  from the upper part of the system. When it is not, the solution is not unique. The programs signal the non-uniqueness and deliver one solution, in which the  $\eta$ 's are set to zero in the upper block of this system.



same sequence of linear operations as applied in the earlier paper to a linear system<sup>7</sup>, we can transform (6) to

$$(13) \quad dy_{it} = G_{1ij} dx_{jt} + G_{2ij} dv_{j,t-1} + G_{3ij} \varepsilon_{jt} + \frac{1}{2} \left( G_{11ijk} dv_{jt} dv_{kt} + 2G_{12ijk} dv_{jt} dv_{k,t-1} \right. \\ \left. + 2G_{13ijk} dv_{jt} \varepsilon_{kt} + G_{22ijk} dv_{j,t-1} dv_{k,t-1} + 2G_{23ijk} dv_{j,t-1} \varepsilon_{kt} + G_{33ijk} \varepsilon_{jt} \varepsilon_{kt} \right)$$

$$(14) \quad J_{1ij} dx_{jt} = J_{2ij} dx_{j,t-1} + J_{3ij} \varepsilon_{jt} + \Pi^* \eta_t + \frac{1}{2} \left( J_{11ijk} dv_{jt} dv_{kt} + 2J_{12ijk} dv_{jt} dv_{k,t-1} \right. \\ \left. + 2J_{13ijk} dv_{jt} \varepsilon_{kt} + J_{22ijk} dv_{j,t-1} dv_{k,t-1} + 2J_{23ijk} dv_{j,t-1} \varepsilon_{kt} + J_{33ijk} \varepsilon_{jt} \varepsilon_{kt} \right),$$

where  $v_t = (y'_t \ x'_t)'$ , i.e. the  $y$  and  $x$  vectors stacked up.

Now the  $y$  and  $x$  introduced above may seem to have no connection to the  $y$  and  $x$  in terms of which we wrote the solution (11)-(12). But that solution has second-order expansion

$$(15) \quad dy_{it} = F_{1ij} dv_{j,t-1} + F_{2ij} \varepsilon_{jt} + F_{3i} \sigma^2 \\ + \frac{1}{2} \left( F_{11ijk} dv_{j,t-1} dv_{k,t-1} + 2F_{12ijk} dv_{j,t-1} \varepsilon_{kt} + F_{22ijk} \varepsilon_{jt} \varepsilon_{kt} \right)$$

$$(16) \quad dx_{it} = \frac{1}{2} M_{11ijk} dy_{jt} dy_{kt} + M_2 \sigma^2.$$

Of course if  $x$  were chosen as an arbitrary linear combination of  $w$ 's, there would in general be a first-order term in  $dy_t$  on the right-hand side of (16). However, we can always move such terms to the left-hand side and then redefine  $x$  to include them. We will now proceed to show that the  $dy$  and  $dx$  in (15)-(16) are indeed those in (13)-(14), and that indeed we can construct the coefficient matrices in the former from knowledge of the coefficient matrices in the latter.

The terms in  $\sigma$  in (15)-(16) deserve discussion. As can be seen from (8), the appearance of expectations operators in our system makes it depend on the distribution of  $\varepsilon$ , not just on realized values of  $\varepsilon$ . But there is only one term in (8) that is first-order in  $dw_{t+1}$ . All the other terms are second-order, or depend on  $dw_t$  or  $\sigma^2$ , not  $\sigma$ . Therefore if there were a component of  $Q_1 K_1 dw_{t+1}$  that depended on  $\sigma$  (rather than  $\sigma^2$ ), that term could not be zero as the equation requires. Hence we can be sure that there is no term linear in  $\sigma$  in

<sup>7</sup>and implemented in the Matlab function `gensys.m`

the second order expansion of (2)-(3), and thus none in (15)-(16). This then also rules out any term of the form  $\sigma \cdot \sigma \varepsilon_{t+1}$  also, since such a term could enter only through the cross products in  $dw_{t+1} \sigma \varepsilon_{t+1}$  or through the  $dw_{t+1} dw_{t+1}$  terms, and without a first-order term in  $\sigma$  in  $dw_{t+1}$ , these cross products can generate no  $\sigma \cdot \sigma \varepsilon_{t+1}$  terms.

Observe that  $dx_t$  in (13)-(14) must be zero to first order (except for  $t = -1$ ), because otherwise there would be an explosive component in the first order part of the solution, contradicting the stability assumption. Therefore,  $F_1$  is exactly  $G_2$  from (13). Clearly also  $F_2 = G_3$ . Therefore we have a complete first-order solution for  $dy$  and  $dx$  in hand:

$$(17) \quad dy_t \doteq F_1 dv_{t-1} + F_2 \varepsilon_t$$

$$(18) \quad dx_t \doteq 0.$$

We find the second order terms in the following steps. First shift (14) forward in time by one (so that the left-hand side is  $dx_{t+1}$ ) and substitute the right-hand side of (16), shifted forward in time by 1, for the  $dx_{t+1}$  on the left. Then substitute the right-hand-side of (17), shifted forward by 1, for all occurrences of  $dy_{t+1}$  in the resulting system. Finally apply the  $E_t$  operator to the result. In doing this, we are dropping all the second order terms in the solution for  $dy$  and  $dx$  when these terms themselves occur in second order terms. This makes sense because cross products involving terms higher than first order are third order or higher, and thus do not contribute to the second order expansion. Note that this means that, since  $dx$  is zero to first order, in (13)-(14) all the second-order terms in  $dv$  can be written in terms of  $dy$  alone. We will abuse notation by using the same  $G$  and  $J$  labels for the smaller second-order coefficient matrices that apply to  $dy$  alone that we use in (13)-(14)

for the second order terms involving the full  $v$  vector. In this way we arrive at

$$\begin{aligned}
(19) \quad J_{1ij} & \left( \frac{1}{2} (M_{11jkl} F_{1kr} dy_{rt} F_{1ls} dy_{st} + M_{11jkl} F_{2kr} F_{2ls} \Omega_{rs} \sigma^2) + M_{2j} \sigma^2 \right) \\
& = J_{2ij} \left( \frac{1}{2} M_{11jkl} dy_{kt} dy_{lt} + M_{2j} \sigma^2 \right) + \frac{1}{2} \left( J_{11ijk} (F_{1jr} F_{1ks} dy_{rt} dy_{st} \right. \\
& \quad + F_{2jr} F_{2ks} \Omega_{rs} \sigma^2) + 2J_{12ijk} F_{1jr} dy_{rt} dy_{kt} + 2J_{13ijk} F_{2jr} \Omega_{rk} \sigma^2 \\
& \quad \left. + J_{22ijk} dy_{jt} dy_{kt} + J_{33ijk} \Omega_{jk} \sigma^2 \right),
\end{aligned}$$

Where we have set  $\text{Var}(\varepsilon_t) = \sigma^2 \Omega$ .

For this equation to hold for all  $dy$  and  $\sigma^2$  values, we must match coefficients on common terms. Therefore, looking at the  $dy_t \cdot dy_t$  terms, we conclude that

$$(20) \quad J_{1ij} M_{11jkt} F_{1kr} F_{1ls} = J_{2ij} M_{11jrs} + J_{11ijk} F_{1jr} F_{1ks} + 2J_{12ijs} F_{1jr} + J_{22ijk}.$$

This is a linear equation, and every element of it is known except for  $M_{11\dots}$ . The transformations that produced the block-recursive system with ordered roots guarantee that  $J_{2\dots}$ , an ordinary  $2 \times 2$  matrix, has all its eigenvalues above the critical stability value. It is therefore invertible, and we can multiply (20) through on the left by  $J_2^{-1}$ , to get a system in the form

$$(21) \quad AM^* F_1 \otimes F_1 = M^* + B.$$

In this equation,  $M^*$  is the ordinary  $n_s \times n_s^2$  matrix obtained by stacking up the second and third dimensions of  $M_{11\dots}$ ,  $A = J_2^{-1} J_1$ , and  $B$  is everything else in the equation that doesn't depend on  $M^*$ . If the dividing line we have specified between stable and unstable roots is  $1 + \delta$ , then our construction of the block-recursive system has guaranteed that  $J_2^{-1} J_1$  has all its eigenvalues  $\leq 1/(1 + \delta)$ , while at the same time it is a condition on the solution that all the eigenvalues of  $F_1$  be  $< 1 + \delta$ . To guarantee that a second-order solution exists, we require that the product of the largest eigenvalue of  $F_1 \otimes F_1$ , which is the square of the largest eigenvalue of  $F_1$ , be less than the inverse of the largest eigenvalue of  $A = J_2^{-1} J_1$ . If  $\delta = 0$  this condition is automatically satisfied. Otherwise, there is an extra condition that was not required for finding a solution to the linear system: the smallest unstable root must exceed the square of the largest stable root.

Assuming this condition holds, (21) has the form of a discrete Lyapunov or Sylvester equation that is guaranteed to have a solution. Because of the special structure of  $F_1 \otimes F_1$ , it would be very inefficient to solve this system with standard packages (like Matlab's `lyap.m`), but it is easy to exploit the special structure with a doubling algorithm to obtain an efficient solution for  $M^*$ .

With  $M_{11\dots}$  in hand, it is easy to see from (19) that we can obtain a solution for  $M_2$  by matching coefficients on  $\sigma^2$ . The only slightly demanding calculation is a required inversion of  $J_2 - J_1$ . But since  $J_2^{-1}J_1$  has all its eigenvalues less than one, this  $J_2 - J_1$  is guaranteed to be nonsingular.

The next step is to use (16) to substitute for the first-order term in  $dx_t$  on the right of (13) and (17)-(18) to substitute for all occurrences of  $dy_t$  and  $dx_t$  in second-order terms on the right in the resulting equation. This produces an equation with  $dy_t$  on the left, and first and second-order terms in  $dy_{t-1}$  and  $\varepsilon_t$  and terms in  $\sigma^2$  on the right. With  $M_{11}$  and  $M_2$  in hand, it turns out that it is only a matter of bookkeeping to read off the values of  $F_{12}$ ,  $F_{22}$ , and  $F_3$  by matching them to the collected coefficients in this equation.<sup>8</sup>

## 5. ANALYZING THE STATE REPRESENTATION

The `gensys.m` program produces as output, among other things, a first-order expansion of (9), as

$$(22) \quad dw_t \doteq F_1^* dw_{t-1} + F_2^* \sigma \varepsilon_t.$$

To find a conventional state-space representation of such a system, we can form a singular value decomposition

$$(23) \quad [F_1^* F_2^*] = [U \quad V] \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} R'_1 & R'_2 \\ S'_1 & S'_2 \end{bmatrix},$$

---

<sup>8</sup>This bookkeeping is not trivial to program, but it is probably best for those who need to program it to consult the program, rather than take up space here with the bookkeeping.

where  $[U \ V]$  and  $[R \ S]$  are orthonormal matrices and  $D$  is diagonal. Any state vector  $z_t$  that has the property that  $w_t$  is determined by  $z_t$  in this system will have to be of the form  $z_t = \theta U' w_t$ . The only way  $w_{t-1}$  affects current  $w_t$  is via  $R'_1 w_{t-1}$ . While  $R'_1$  can be the same row rank as  $U$ , it can also be less, so that a smaller “state” vector summarizes the past than is needed to characterize the current situation. Also, the rank of  $F_1^*$  can be below its number of non-zero singular values. In this case it may be possible to find a  $z_t$  that, after the system has run a few periods, summarizes the past and/or characterizes the current situation yet is of lower dimension than the rank of  $D$ .

The program `gstate.m` takes as input  $F_1^*$  and  $F_2^*$ , together with an optional candidate matrix  $\phi$  of coefficients that might form a state vector as  $z_t = \phi w_t$ . The program checks whether  $\phi$  lies in  $U$ 's or  $R'_1$ 's row space and returns  $U$  and  $R'_1$  for further analysis.

Once a state/co-state representation of the form  $v_t = \Psi w_t$  has been settled on, where  $\Psi$  is non-singular and the  $v = (y' \ x)'$  vector is partitioned into state and costate, it is straightforward to convert a first or second-order approximate solution from one co-ordinate system into the another.

## 6. THE LOCAL ACCURACY OF THE APPROXIMATION

Once we have a second-order accurate approximation to the dynamics, in the form (15)-(16), we can make a claim to local accuracy of the following form:

$$(24) \quad dv_{t+1} = \hat{F}(dv_t, \sigma \varepsilon_{t+1}, \sigma) + o_p(\|dv_t, \sigma\|^2),$$

where  $o_p$  means “order in probability” and  $\hat{F}$  is the second-order approximation to the dynamics. That is, the error in the approximation is claimed to converge in probability to zero, at a more rapid rate than  $\|dv_t, \sigma\|^2$ , when  $\|dv_t, \sigma\|^2$  goes to zero. This rate is the weakest kind of claim that can be made for a Taylor expansion. If we are willing to claim that third derivatives exist at the deterministic steady state, then we can replace the error term with  $O_p(\|dv_t, \sigma\|^3)$ . This claim does not depend on strict boundedness of the support of the distribution of  $\varepsilon_t$ , because we are only claiming our local accuracy with a certain (high) probability. Whatever the distribution of  $\varepsilon_t$ ,  $\sigma \varepsilon_t$  converges in probability to zero as

$\sigma \rightarrow 0$ , allowing us to make this claim. Of course this is all dependent on the underlying assumption that the original nonlinear model has dynamics differentiable of sufficiently high order in  $\sigma$  in the neighborhood of deterministic steady state, and on the existence and continuity of the expectations that occur in the statement of the model.

This one-step-ahead “local accuracy in probability” claim obviously can be extended to a corresponding claim to accuracy  $n$ -steps-ahead for any finite  $n$ . We have made no appeal to stationarity of the system in making these claims. Of course the size of the  $n$  for which accuracy remains good at a given level of  $\sigma$  will in general be smaller for systems that are not stationary. But the qualitative nature of the accuracy claim is no different for non-stationary systems.

This type of finite-time-span, accuracy-in-probability claim is exactly what is appropriate for purposes of fitting a model to data — which always cover a finite time span — or for purposes of simulating the model from given initial conditions over a finite span of time. It is also exactly appropriate for the correct calculation of expected welfare, when welfare is constructed as a discounted sum of period utilities. The discounting means that accuracy of the approximation is unimportant after some time horizon in the future.

Finite-time-span, accuracy-in-probability claims will not justify estimating unconditional expectations of any functions of variables in the model via simulation. To make the effects of initial conditions die away, such simulations must cover long spans of time. If the second-order approximation is non-stationary, expectations calculated from simulations of it will of course not converge. If the true nonlinear model is non-stationary, then the true unconditional expectations will in general not exist, even though it is possible that the local second-order approximation is stationary, so again in this case it will not be possible to estimate unconditional expectations from simulated paths.

When both the true nonlinear model and the second order approximate model are stationary and ergodic, and the true unconditional expectation in question is a twice-differentiable function of  $\sigma$  in the neighborhood of  $\sigma = 0$ , then it is possible to estimate the expectation from long simulations of the approximate model, with the estimates accurate locally in  $\sigma$

in the usual sense. This is true even though it may be (e.g. because of unbounded support of  $\varepsilon_t$ ) that with probability one the path of the model repeatedly enters regions where the local approximation is inaccurate. This is possible because as  $\sigma \rightarrow 0$  the fraction of time spent in these regions goes to zero, for both the true and the approximate model.

However it will often be preferable to estimate an expectation by using the second-order approximation analytically, expanding the function whose expectation is being taken as a Taylor series and applying the methods of the next section.

It goes without saying that no theoretical result about local or asymptotic global accuracy for approximate solutions can prove that in a particular model, with particular shock variances, one method or another is more accurate than another or accurate enough for some specific purpose. The emphasis by Jin and Judd (2002) on checking model validity is therefore appropriate. There is no uniquely best measure of solution accuracy, but by now a variety of stringent checks have been proposed. The methods that have been applied most widely (but not widely enough) are based on evaluating the conditional expectation in (3) at a collection of values of the lagged state vector. There are important practical questions as to how to select the collection of state vector values at which one evaluates the expectation and as to what metric to use in measuring the vector of deviations from the theoretical zero values for these expectations. Jin and Judd suggest deterministically fixing a collection of state variable values and a set of “relative error” metrics for the expectational errors, based on economic interpretation of the model being solved. den Haan and Marcet (1994) suggest another approach, in which the state variable values are generated stochastically via simulation and the metric for evaluation of expectational errors is based on statistical detectability of the errors in a sample of relevant length. Each of these approaches has pitfalls, but is worth consideration.

Though there is no widely understood alternative to this “Euler equation residual” family of accuracy checks at this point, there is probably room for further work in this area. For many purposes, the most relevant measure of accuracy is the accuracy of the solution’s approximation to the mapping from  $w_{t-1}$ ,  $\varepsilon_t$ , and  $\sigma$  to  $w_t$  corresponding to (9). This is not

measured directly by the size of the Euler equation errors, but no more direct measure of the accuracy of this mapping is at this point available.

## 7. FORECASTING AND SIMULATION

Forecasts  $s$  steps ahead,  $E_t[dw_{t+s}]$  and  $\text{Var}_t[dw_{t+s}]$  are the building blocks for the calculation of impulse response functions as well as welfare.

We build the forecasts from the second-order accurate dynamic model given by (15)-(16), modified here to reflect our assumption that the initial conditions satisfy the equations of the model and that therefore  $dx_{t-1} = 0$  to first order. We abuse notation by using the same  $F$ 's here, for the pieces of the original  $F$  matrices corresponding to  $dy$ 's, as we did for the original  $F$  matrices in (15)-(16) that corresponded to the full  $dv = [dy, dx]$  vector.

$$(25) \quad dy_t \doteq F_{1j}dy_{j,t-1} + F_{2j}\sigma\varepsilon_{j,t} + F_3\sigma^2 + \frac{1}{2}F_{11jk}dy_{j,t-1}dy_{k,t-1} + F_{12jk}\sigma dy_{j,t-1}\varepsilon_{k,t} + \frac{1}{2}F_{22jk}\sigma^2\varepsilon_{j,t}\varepsilon_{k,t}$$

$$(26) \quad dx_t \doteq \frac{1}{2}M_{11jk}dy_{j,t}dy_{k,t} + M_2\sigma^2$$

We would then like to calculate, to second order accuracy,  $E_t[dy_{t+s}]$  and  $\text{Var}_t[y_{t+s}]$ .

To begin with, note that, since the conditional mean of  $dy_{t+s}$  is of second order, the variance terms  $\hat{\Sigma}_s \equiv \text{Var}_t(y_{t+s})$  are correct to second order accuracy when computed from the first-order terms in the expansion (25) alone and that, to second-order accuracy,  $\text{Var}_t(x_{t+s}) = 0$  since  $dx_t$  itself is of second order.

For  $s = 1$ , it is easy to see from (25)-(26) that we have

$$(27) \quad d\hat{y}_{t+1} = E_t[dy_{t+1}] \doteq F_{1j}dy_{j,t} + F_3\sigma^2 + \frac{1}{2}F_{11jk}dy_{jt}dy_{kt} + \frac{\sigma^2}{2}F_{22jk}\Omega_{jk}$$

$$(28) \quad d\hat{x}_{t+1} = E_t[dx_{t+1}] \doteq \frac{1}{2}M_{11jk}(d\hat{y}_{j,t+1}d\hat{y}_{k,t+1} + \hat{\Sigma}_{1,jk}) + M_2\sigma^2.$$

The expression in (28) for determining  $E_t[dx_{t+1}]$  from the conditional mean and variance of  $dy_{t+1}$  works equally well for determining  $E_t[dx_{t+s}]$  from the conditional mean and variance of  $dy_{t+s}$  for  $s > 1$ . The straightforward approach to determining  $d\hat{y}_{t+s}$  and  $d\hat{x}_{t+s}$  is to apply



(27) recursively, computing  $d\hat{y}_{t+s}$  from  $d\hat{y}_{t+s-1}$  and  $\hat{\Sigma}_{k-1}$ , etc. This procedure is in fact second-order accurate, but it introduces higher order terms into the expansion. For example, since  $\hat{d}y_{t+1}$  contains quadratic terms in  $dy_t$ , and (27) makes  $d\hat{y}_{t+2}$  quadratic in  $d\hat{y}_{t+1}$ , in a simple recursive computation  $d\hat{y}_{t+2}$  becomes quartic in  $dy_t$ . These extra high-order terms do not in general increase accuracy of the approximation, as they do not correspond to higher order coefficients in a Taylor series expansion of the true dynamic system, and in practice often lead to explosive time paths for  $d\hat{y}_{t+s}$ .

To see what goes wrong, consider the simple univariate model

$$y_t = \rho y_{t-1} + \alpha y_{t-1}^2 + \varepsilon_t,$$

where  $|\rho| < 1$  and  $\alpha > 0$ . Though this model is locally stable about its unique deterministic steady state of  $\bar{y} = 0$ , it has a second steady-state, at  $(1 - \rho)/\alpha$ . If  $x$  exceeds the other steady state, it will tend to diverge. This is likely to be a generic problem with quadratic expansions — they will have extra steady states not present in the original model, and some of these steady states are likely to mark transitions to unstable behavior.

Since the unique local dynamics are stable in a neighborhood of the steady state, it will be desirable to choose amongst the second order accurate expansions one that implies stability. Deriving sufficient conditions on the support of  $\varepsilon_t$  to guarantee non explosiveness under the iterative scheme (27)-(28) is in general a non-trivial task and therefore it is useful to have available an algorithm which generates non-explosive forecasts and simulations without imposing explicit conditions on the support of  $\varepsilon_t$ . The mere fact that the generated forecasts are stable of course does not imply superior accuracy in general, especially when shocks are not bounded. However, stationarity will in general imply that, for a given neighborhood  $\mathcal{U}$  of the steady state and a given time horizon  $T$ , we can restrict  $\sigma$  in such a way as to make the probability of leaving  $\mathcal{U}$  in time  $T$  arbitrarily small.

Obtaining a stable solution based on (27) can be achieved by pruning out the extraneous high-order terms in each iteration by computing the projections of the second order terms based on a *first-order* expansion,  $d\bar{y}_{t+1}$  of  $E_t[dy_{t+1}]$ , as follows:

$$(29) \quad d\hat{y}_{t+s} \doteq F_{1j}d\hat{y}_{j,t+s-1} + F_3\sigma^2 \\ + \frac{1}{2}F_{11jk} (d\bar{y}_{j,t+s-1}d\bar{y}_{k,t+s-1} + \hat{\Sigma}_{k-1,jk}) + \frac{\sigma^2}{2}F_{22jk}\Omega_{jk}$$

$$(30) \quad d\hat{x}_{t+s} \doteq \frac{1}{2}M_{11jk} (d\bar{y}_{j,t+s}d\bar{y}_{k,t+s} + \hat{\Sigma}_{s,jk}) + M_2\sigma^2$$

$$(31) \quad d\bar{y}_{t+s} \doteq F_{1j}d\bar{y}_{j,t+s-1}$$

$$(32) \quad \hat{\Sigma}_{i,j,s} = \sigma^2 F_{2ik}\Omega_{kl}F_{2jl} + F_{1ik}\hat{\Sigma}_{k\ell,s-1}F_{1j\ell}.$$

Using these equations recursively results in a  $d\hat{y}_{t+s}$  series which, by construction, is quadratic in  $dy_t$  for all  $s$ . Furthermore, when the eigenvalues of  $F_1$  are less than one in absolute value, the first order accurate solution  $d\bar{y}_{t+s}$  is stable and hence so is the squared process  $(d\bar{y}_{j,t+s}d\bar{y}_{k,t+s})$ . It follows that  $d\hat{y}_t$  must be stable as well.<sup>9</sup> Note that the  $F_{12}$  component of the second order expansion — the coefficients of the interactions between  $dy_{t-1}$  and  $\varepsilon_t$  — do not enter this recursion at all.

The same issues arise if the aim is to generate simulated time paths, rather than simply conditional expectations and variances of future variables. For this purpose, we can introduce the notation  $dy_{t+s}^{(1)}$  and  $dy_{t+s}^{(2)}$  for first and second order accurate simulated time paths, respectively. A recursive, non-explosive, “pruned” simulation scheme is then given by

$$(33) \quad dy_{t+s}^{(2)} \doteq F_{1\cdot j}dy_{j,t+s-1}^{(2)} + F_{2\cdot j}\sigma\varepsilon_{j,t+s} + F_3\sigma^2 \\ + \frac{1}{2}F_{11\cdot jk}dy_{j,t+s-1}^{(1)}dy_{k,t+s-1}^{(1)} + \sigma F_{12\cdot jk}y_{j,t+s-1}^{(1)}\varepsilon_{k,t+s} + \frac{\sigma^2}{2}F_{22\cdot jk}\varepsilon_{j,t+s}\varepsilon_{k,t+s}$$

$$(34) \quad dx_{t+s}^{(2)} \doteq \frac{1}{2}M_{11\cdot jk}dy_{j,t+s}^{(1)}dy_{k,t+s}^{(1)} + M_2\sigma^2$$

$$(35) \quad dy_{t+s}^{(1)} \doteq F_{1\cdot j}dy_{j,t+s-1}^{(1)} + F_{2\cdot j}\sigma\varepsilon_{j,t+s},$$

where the  $F_{12}$  terms that could be ignored in forming conditional expectations have necessarily returned for generation of accurate simulations. By preventing buildup of spurious higher-order terms, we make stability of the simulation over a long time path more likely,

<sup>9</sup>The same matrix eigenvalue conditions are at issue here as in section 4’s discussion of existence of the solution to (21)

while at the same time preserving second-order accuracy of the mapping from initial variable values  $y_t, x_t$ , shocks  $\varepsilon_{t+1}, \dots, \varepsilon_{t+s}$ , and  $\sigma$  to the simulated values  $y_{t+1}^{(2)}, \dots, y_{t+s}^{(2)}$ .

It can help in understanding these recursions to append the vector  $dy^{(1)} \otimes dy^{(1)}$  to  $dy^{(2)}$  and use matrix notation:

$$(36) \quad \begin{bmatrix} dy_{t+1}^{(2)} \\ (dy_{t+1}^{(1)} \otimes dy_{t+1}^{(1)}) \end{bmatrix} = \Theta_1 \begin{bmatrix} dy_t^{(2)} \\ (dy_t^{(1)} \otimes dy_t^{(1)}) \end{bmatrix} + \Theta_2 \sigma^2 + \xi_{t+1}$$

with

$$(37) \quad \Theta_1 = \begin{bmatrix} F_1 & \frac{1}{2}(F_{11}^*) \\ 0 & (F_1 \otimes F_1) \end{bmatrix}$$

$$(38) \quad \Theta_2 = \begin{bmatrix} F_3 + \frac{1}{2}F_{22 \cdot jk} \Omega_{jk} \\ (F_2 \otimes F_2) \text{vec}(\Omega) \end{bmatrix}$$

$$(39) \quad \xi_t = \begin{bmatrix} \sigma \varepsilon_t + \sigma F_{23 \cdot jk} \varepsilon_{k,t} dy_{j,t-1}^{(1)} + \frac{\sigma^2}{2} F_{33 \cdot jk} (\varepsilon_{jt} \varepsilon_{kt} - \Omega_{jk}) \\ \sigma^2 (\varepsilon_t \otimes \varepsilon_t - (F_2 \otimes F_2) \text{vec}(\Omega)) \end{bmatrix}.$$

The  $F_{11}^*$  in the definition of  $\Theta_1$  (37) is a matrix with number of rows equal to the length of  $y$  and with the second and third dimensions of the array vectorized into a row vector — so it is an  $n_s \times n_s^2$  matrix. Note that  $\Theta_1$  is upper block triangular and is stable exactly when the eigenvalues of  $F_1$  are less than one in absolute value. Note also that, to second order accuracy,

$$\text{Var}(\xi_t) = \begin{bmatrix} \sigma^2 \Omega & 0 \\ 0 & 0 \end{bmatrix}.$$

Calculations of conditional and unconditional first and second moments can therefore be carried out using (36) as if it were an ordinary first order VAR. This can be an aid to understanding, or to computation in small models, though for larger systems it is likely to be important for computational efficiency to take account of the special structure of the  $\Theta_j$  matrices in (36).

## 8. WELFARE

One can easily produce cases where the second-order approximation is necessary to get an accurate evaluation of certain aspects of the model. Utility-based welfare calculation is one case. For example, calculating welfare effects of various monetary and fiscal policies or welfare effects of changes in economic environment such as financial market structure should include second-order or even higher-order terms in order to get an accurate measure. Kim and Kim (2003a) present an example of how inaccurate the linearized solution can be in calculating welfare using a two-country model. Using the linearized solution, welfare of autarky can appear to be higher than that of the complete markets, solely because of the inaccuracy of the linearization method. Another application in which second-order approximation is important is examination of asset price behavior in DSGE's. Linearized solutions will imply equal expected returns on all assets. Second order solutions will generate correct risk premia, though generally to analyze time variation in risk premia will require higher than second-order accuracy.

Equation (36) makes it relatively straightforward to see how to carry out a second-order accurate welfare calculation. Welfare is defined as a discounted sum of expected utility. Let the period utility function be given by  $u : \mathbb{R}^{n_s} \rightarrow \mathbb{R}$ .<sup>10</sup> Then the utility conditional on an

---

<sup>10</sup>Of course often in growth models utility is a function of consumption, which is not a conventional state variable. To use the formulation we develop here, then, consumption (an  $x$  variable) has to be replaced by the corresponding component of  $h(y, \sigma)$ . Also, because we work entirely in terms of  $y$ , we are not covering the case where the initial distribution of  $w$  does not lie on the saddle path. The methods we describe here can be expanded to cover this case and to allow  $x$  to enter  $u$ , at the cost of some increase in the burden of notation.

initial distribution of  $y_0$  with mean and variance  $(\mu, \Sigma)$  is

$$\begin{aligned}
 (40) \quad U(\mu, \Sigma) &= E_0 \left[ \sum_{t=0}^{\infty} \beta^t u(y_t) \right] \approx \\
 &\frac{u(\bar{y})}{1-\beta} + E_0 \left[ \sum_{t=0}^{\infty} \beta^t (\nabla u(\bar{y}) dy_t^{(2)} + \frac{1}{2} \text{vec}(\nabla^2 u(\bar{y}))' (dy_t^{(1)} \otimes dy_t^{(1)})) \right] \Rightarrow \\
 (41) \quad U(\mu, \Sigma) &= \frac{u(\bar{y})}{1-\beta} + \left[ \nabla u(\bar{y}) \quad \frac{1}{2} \text{vec}(\nabla^2 u(\bar{y}))' \right] \\
 &\cdot [I - \beta \Theta_1]^{-1} \left( \begin{bmatrix} \mu \\ \text{vec}(\Sigma + \mu \mu') \end{bmatrix} + \beta(1-\beta)^{-1} \Theta_2 \sigma^2 \right)
 \end{aligned}$$

If we are interested only in unconditional expected  $u$ , we can arrive at the correct formula by multiplying (41) through by  $1 - \beta$  and taking the limit as  $\beta \rightarrow 1$ , giving us

$$(42) \quad E[u(y_t)] = u(\bar{y}) + \left[ \nabla u(\bar{y}) \quad \frac{1}{2} \text{vec}(\nabla^2 u(\bar{y}))' \right] (I - \Theta_1)^{-1} \Theta_2 \sigma^2.$$

Note that in (41) we make no use, explicitly or implicitly, of  $F_{12}$ . Also note that though the matrix  $I - \beta \Theta_1$  appears in the formula inverted, the utility calculation only requires

$$\left[ \nabla u(\bar{y}) \quad \frac{1}{2} \text{vec}(\nabla^2 u(\bar{y}))' \right] \cdot (I - \beta \Theta_1)^{-1},$$

whose computation is only an equation-solving problem, not a full inversion;<sup>11</sup> furthermore, this part of the computation does not need to be repeated as  $\mu$  and  $\Sigma$  are varied. Finally, note that (42) uses only  $(I - \Theta_1)^{-1} \Theta_2$ , regardless of the form of  $u$ . This is again an equation-solving problem. So if we are interested only in unconditional expectations, even in unconditional expectations of many different functions  $u$ , the computation of a full second-order correction may be much simpler than calculation of the full second-order expansion of the dynamics.

<sup>11</sup>Though for an  $n \times n$  matrix  $A$  both solving  $Ax = b$  for  $x$  and computing  $A^{-1}$  are  $O(n^3)$  operations, the latter is substantially more time consuming. In Matlab inversion takes roughly twice the time.

It is these simplifications, applied to particular models, that are the insights provided by the papers that have put forward “bias-correction” methods for making second-order accurate expected welfare computations in DSGE models (Kim and Kim, 2003a; Sutherland, 2002).

We should note that there is a situation in which second-order accurate evaluations of welfare can avoid entirely the need for a second-order expansion of the model solution. If  $\nabla u(\bar{y}) = 0$ , as would be true if the deterministic steady-state sets  $y$  to the value that maximizes  $u(y)$ , then only the lower blocks of  $\Theta_1$  and  $\Theta_2$  enter the solution, as can be seen from (41) or (42). As can be seen from (37) and (39), these blocks contain  $F_1$  and  $F_2$  only, not any terms from the second-order solution. Of course in most problems with discounting, even an optimal solution will not maximize static welfare  $u(y)$  in the steady state, so this result will not apply. Also, even where the solution has been computed to maximize static period welfare  $u$ , the result depends on having a second order expansion of  $u$  in terms of the state vector  $y$ . When the problem has been formulated (as in usual growth models) with a non-state variable (e.g. consumption) appearing in the utility function, the second-order expansion of the utility function in terms of  $y$  may require use of the second-order solution for  $x$  as a function of  $y$ .<sup>12</sup>

**8.1. Conditional vs. Unconditional welfare.** From the discussion in the preceding section it is apparent that evaluating expected welfare based on unconditional  $E[u(y)]$  is a more straightforward task than evaluating the conditional expectation of discounted expected utility at a given date.<sup>13</sup> It is therefore not surprising that many existing papers have

<sup>12</sup>Rotemberg and Woodford (1997) is an example of a context where use of the first-order solution for welfare analysis is justified by special regularity conditions. The paper evaluated welfare using unconditional expectation of period utility. Regularity conditions required to justify use of the first-order solution in the paper’s model include an assumption that some other policy change perfectly offsets second-order effects of monetary policy on the mean level of output and an assumption that monetary policy is the only source of inefficient fluctuations in prices.

<sup>13</sup>Woodford (2002) discusses the differences between unconditional and conditional welfare in calculating welfare effects of monetary policies.

used unconditional welfare for evaluating policies. Examples include Clarida, Galí, and Gertler (1999), Rotemberg and Woodford (1997, 1999), Sutherland (2002) and Kollmann (2002).

There are strong objections in principle to use of the unconditional welfare criterion. We know that it takes time for one steady state to reach another steady state and unconditional welfare neglects the welfare effects during the transitional period. It is therefore generally not in fact optimal, in problems with discounting, to use policies that maximize the unconditional expectation of one-period welfare. This is not a new point — it is the same point as the non-optimality of driving the rate of return to zero in a growth model — and it has been recognized in the DSGE literature in, e.g, Kim and Kim (2003b), and Woodford (2002).

Because unconditional welfare can often be computed easily, using the “bias correction” shortcut, it is important to note that using unconditional welfare can give nonsensical results. Kim and Kim (2003a) construct a two-country DSGE model and compute risk-sharing gains from autarky to the complete-markets economy using a second-order approximation method. Welfare is defined as conditional welfare and the results show that there are positive welfare gains from autarky to the complete-markets economy. But the unconditional welfare measure can for certain parameter values produce the paradoxical result that autarky generates a higher level of welfare than the complete markets.

The use of conditional welfare does not imply that results necessarily are tied to some particular initial state. One can condition on a distribution of values for the initial state. The critical point is that when comparing two policies or equilibria one should use the same distribution for the initial state for each. When there is no time-inconsistency problem the optimal policy will have the property that no matter what initial distribution is specified for the state, it will produce a higher conditional expectation of welfare than any other policy. However, when comparing a collection of policies that are not optimal, one may find that rankings of policies vary with the assumed distribution of the initial state.

When there is a time-inconsistency problem, the optimal policy generally depends on the initial conditions, even if we restrict attention to policy rules that are a fixed mapping from

state to actions. Using a conditional expectation as the welfare measure does not avoid this problem. One attempt to get around this issue is the suggestion in, e.g., Giannoni and Woodford (2002) that policy should follow the rule that would prevail under commitment in the limit as the initial conditions recede into the past. This “timeless perspective” policy can be implemented by treating the Lagrange multipliers on private sector Euler equations as “states”, and then maximizing conditional expected discounted utility. The timeless perspective policy is a useful benchmark, but it cannot resolve the fundamental problem of time inconsistency. A policy-maker who can make believable commitments will not want to choose the timeless-perspective solution, while one that cannot make believable commitments cannot implement the timeless-perspective solution. As a normative suggestion, the timeless perspective depends on the idea that it will be easier to convince the public of a commitment to the timeless-perspective policy than of commitments to other types of policies, but this is likely to depend on the nature of the policy optimization problem and on the particular initial conditions faced at the time the policy is implemented.

## 9. CONCLUSION

Use of perturbation methods to improve analysis of DSGE models is still in its early stages. Programs that automate computations for models higher than second order are just beginning to emerge. Methods of dealing with the kinds of singularities that show up in economic models — for example the indeterminacy of asset allocations in standard portfolio problems when variances are zero — are still not widely understood. And we have only begun to get a feel for where these methods are useful and what their limitations are. Real progress is being made, however, in an atmosphere that is both competitive enough to be stimulating and cooperative enough that researchers located around the world are benefiting from each others’ insights.



## REFERENCES

- ANDERSON, G., AND A. LEVIN (2002): “A user-friendly, computationally-efficient algorithm for obtaining higher-order approximations of non-linear rational expectations models,” Discussion paper, Board of Governors of the Federal Reserve.
- BERGIN, P. R., AND I. TCHAKAROV (2002): “Does Exchange Rate Risk Matter for Welfare? A Quantitative Investigation,” Discussion paper, University of California at Davis, <http://www.econ.ucdavis.edu/faculty/bergin/>.
- CLARIDA, R., J. GALÍ, AND M. GERTLER (1999): “The science of monetary policy: A new Keynesian perspective,” *Journal of Economic Literature*, 37, 1661–1707.
- COLLARD, F., AND M. JUILLARD (2000): “Perturbation Methods for Rational Expectations Models,” Discussion paper, CEPREMAP, Paris, [fabrice.collard@cepremap.cnrs.fr](mailto:fabrice.collard@cepremap.cnrs.fr).
- DEN HAAN, W. J., AND A. MARCET (1994): “Accuracy in Simulations,” *The Review of Economic Studies*, 61(1), 3–17.
- GASPAR, J., AND K. L. JUDD (1997): “Solving large-scale rational expectations models,” *Macroeconomic Dynamics*, 1, 44–75.
- GIANNONI, M. P., AND M. WOODFORD (2002): “Optimal Interest-Rate Rules: I. General Theory,” Discussion paper, Columbia University and Princeton University, <http://www.princeton.edu/~woodford>.
- JIN, H., AND K. L. JUDD (2002): “Perturbation Methods for General Dynamic Stochastic Models,” Discussion paper, Stanford University, [judd@hoover.stanford.edu](mailto:judd@hoover.stanford.edu).
- JUDD, K. L. (1998): *Numerical Methods in Economics*. MIT Press, Cambridge, Mass.
- KIM, J., AND S. KIM (2003a): “Spurious Welfare Reversals in International Business Cycle Models,” *Journal of International Economics*, 60, 471–500.
- (2003b): “Welfare Effects of Tax Policy in Open Economies: Stabilization and Cooperation,” Discussion paper, University of Virginia and Tufts University, <http://www.tufts.edu/~skim20>.

- KING, R. G., AND M. WATSON (1998): "The Solution of Singular Linear Difference Systems Under Rational Expectations," *International Economic Review*, 39(4), 1015–1026.
- KLEIN, P. (2000): "Using the generalized Schur form to solve a multivariate linear rational expectations model," *Journal of Economic Dynamics and Control*, 24(10), 1405–1423.
- KOLLMANN, R. (2002): "Monetary Policy Rules in the Open Economy: Effects on Welfare and Business Cycles," *Journal of Monetary Economics*, 49, 989–1015, <http://www1.wiwi.uni-bonn.de/users/rkollmann/www/>.
- ROTEMBERG, J. J., AND M. WOODFORD (1997): "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy," *NBER Macro Annual*, 12, 297–345.
- (1999): "Interest rate rules in an estimated sticky-price model," in *Monetary Policy Rules*, ed. by J. B. Taylor. University of Chicago Press, Chicago.
- SCHMITT-GROHÉ, S., AND M. URIBE (2002): "Solving dynamic general equilibrium models using a second-order approximation to the policy function," Discussion paper, Rutgers University and University of Pennsylvania.
- SIMS, C. A. (2001): "Solving Linear Rational Expectations Models," *Computational Economics*, 20(1-2), 1–20, <http://www.princeton.edu/~sims/>.
- SUTHERLAND, A. (2002): "A simple second-order solution method for dynamic general equilibrium models," Discussion paper, University of St. Andrews, <http://www.st-andrews.ac.uk/~ajs10/home.html>.
- WOODFORD, M. (2002): "Inflation Stabilization and Welfare," *Contributions to Macroeconomics*, 2(1), Article 1.

FEDERAL RESERVE BOARD, TUFTS UNIVERSITY, NORTHWESTERN UNIVERSITY, PRINCETON UNIVERSITY

*E-mail address:* [sims@princeton.edu](mailto:sims@princeton.edu)