NBER WORKING PAPER SERIES

CAUSALITY, STRUCTURE, AND THE UNIQUENESS OF RATIONAL EXPECTATIONS EQUILIBRIA

Bennett T. McCallum

Working Paper 15234
http://www.nber.org/papers/w15234

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2009

The author is indebted to Seonghoon Cho, Robert Lucas, and Holger Sieg for helpful comments on earlier drafts. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2009 by Bennett T. McCallum. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Causality, Structure, and the Uniqueness of Rational Expectations Equilibria
Bennett T. McCallum
NBER Working Paper No. 15234
August 2009
JEL No. C61,C62,E37

ABSTRACT

Consider a rational expectations (RE) model that includes a relationship between variables x_t and z_{t+1}. To be considered structural and potentially useful as a guide to actual behavior, this model must specify whether x_t is influenced by the expectation at t of z_{t+1} or, alternatively, that z_{t+1} is directly influenced (via some inertial mechanism) by x_t (i.e., that z_t is influenced by x_{t-1}). These are quite different phenomena. Here it is shown that, for a very broad class of multivariate linear RE models, distinct causal specifications involving both expectational and inertial influences will be uniquely associated with distinct solutions—which will result operationally from different specifications concerning which of the model's variables are predetermined. It follows that for a given structure, and with a natural continuity assumption, there is only one RE solution that is fully consistent with the model's specification. Furthermore, this solution does not involve "sunspot" phenomena.

Bennett T. McCallum
Tepper School of Business, Posner 256
Carnegie Mellon University
Pittsburgh, PA 15213
and NBER
bm05@andrew.cmu.edu

1. Introduction

It is well known that there exists a huge and still-growing literature concerning the multiplicity of rational expectations (RE) solutions, i.e., processes that satisfy the relevant model's equations and RE orthogonality conditions. In recent monetary economics, the usual response of researchers has been to hope for a unique stable solution while designing policy so as to avoid indeterminacy, the latter being defined as a system with more than one stable solution. There are several ways, however, in which this approach is unsatisfactory. Cochrane (2007), for example, argues that a finding of determinacy—i.e., a single stable solution—is not sufficient to imply a particular inflation outcome. From a different perspective, McCallum (2003, 2007) argues that determinacy is not necessary for a unique solution to be implied because learnability [as developed in Evans and Honkapohja (2001)] is necessary for a solution to be plausible. In addition, Cho and McCallum (2009) describe "another weakness" of determinacy as a selection criterion. Of course, indeterminacy also arises in non-monetary models including the Calvo (1979) overlapping-generations model of land pricing or various examples with increasing returns.

Here I wish to argue that there is an important sense in which these RE solutions reflect not a multiplicity of solutions for a single model, but instead a multiplicity of models each with a single solution. Specifically, I will argue that if attention is paid to the "direction of causality" of intertemporal relationships—expectational vs. inertial—then one of the RE solutions stands out uniquely as a candidate for "equilibrium." In addition, structural models—e.g., ones based on optimizing analysis—typically require such attention in their specification. In making this argument, I will add only one new ingredient to those standard in the RE literature; it constitutes an assumption of continuity of solution coefficients with respect to structural parameters (in the vicinity of zero).

I begin by asserting that there can be little if anything more fundamental and critical, in dynamic economic modelling, than the specification and interpretation of the direction of causality in intertemporal relationships. For example, the "model"

(1) $y_t - \alpha y_{t+1} = 0$

can be interpreted as representing a system in which the variable y_t is determined by agents' expectations in t of y_{t+1} or, alternatively, as representing a system in which y_{t+1} is directly influenced by the previous period's realization of y_t (or, equivalently, that y_t is influenced by y_{t-1}). These two interpretations or specifications represent drastically different models of y_t determination.[1] The first features a crucial role for agents' current expectations of future values (with no influence from the past) whereas the latter assigns an impact to past values via, e.g., adjustment costs, and has no role for expectations (possibly because of an extreme discounting of the future).[2] In terms of dynamic properties, to continue with the contrast, in the former case the system will be dynamically stable for any finite value of the parameter α, while the latter case features dynamic stability only if $|\alpha| > 1$. Also, simulations in stochastic versions with rational expectations are conducted quite differently in the two cases.

Accordingly, specifying the postulated direction of causation for all relations is an essential part of a model's specification, if that model is intended to represent the way in which data is generated by some economic system in which agents' expectations are

[1] It should be clear that we are referring to "causality" in the model-specification sense of Simon (1953), not to Granger causality. For a useful discussion, see Zellner (1979), especially pp. 21-25.
[2] In the first, the causality is not unidirectional; instead y_t and expectations about y_{t+1} are determined jointly. In the second, however, y_{t-1} is not influenced by y_t—it is predetermined.

potentionally important.[3] Of course it is the case that many—perhaps most—models include both expectational and inertial components. (These are the terms that will be used henceforth to refer to the two types of influences.) In such cases, the considerations just described remain fundamental; among other things, it is essential not to confuse parameter values relating to expectational components with those descriptive of adjustment-cost or other inertial aspects of the modelled mechanism. Any model that purports to be structural must surely be clear about all such distinctions. Indeed, all relevant causal specifications will be generated automatically in any model that is based on explicit analysis of agents' optimization problems plus market clearing (as, e.g., in so-called DSGE models).

How is causality specification accomplished, operationally, in the example under discussion, $y_t - \alpha y_{t+1} = 0$? The answer is the same whether or not there is a stochastic component. The direction of causation is determined by specifying whether y_t is or is not predetermined in relation (1)—i.e., unaffected by developments in the period to which the relationship pertains.[4] Indeed, it will be shown in what follows that multiple solutions in linear rational expectations (RE) models invariably reflect multiple specifications regarding which variables are predetermined and which are not, and these specifications are in fact the operational counterpart of causality specifications. As most of my examples will for simplicity exclude exogenous variables, the predetermined variables will in those cases also be the system's state variables.

[3] If the first interpretation above is put forth for model (1), then there is no place for concern regarding dynamic stability. Observation of explosive tendencies in an empirical study of y_t behavior would, accordingly, tend to discredit a hypothesis to the effect that the model with the first interpretation is appropriate for the data-generating mechanism at hand.

[4] If y_t is taken to be predetermined in (1) it might be natural to write the relation as $(1/\alpha)y_t - y_{t+1} = 0$ or, equivalently, $y_t = (1/\alpha)y_{t-1}$.

The purpose of the present paper is to argue that recognition of the importance of causality specification, in the sense just described—i.e., of distinguishing expectational influences from inertial influences via adjustment costs, lags, etc.—will, with the adoption of a simple and natural continuity property, eliminate issues relating to possible "indeterminacy" of the multiple-stable-solution type in linear rational expectations (RE) models.[5] Specifically, there is in each model only one RE solution, which may be dynamically stable or unstable, that accurately reflects a given causality specification—that is, reflects a given specification of which variables in the system are predetermined. This solution might be regarded as representing a proposed equilibrium refinement. The continuity property that will be adopted is that polynomials and eigenvalues relating to solution parameters are continuous functions of the model's structural parameters. Analysis involving these properties has a long and honorable history in economics, physics, and engineering. Some analysts may not be attracted by it, but many, I am confident, will find it both attractive in principle and useful in practice.

2. Basic Univariate Model

Let us begin the discussion with a univariate linear model that features inclusion of both expectational and inertial influences, assuming of course that the analyst specifies which is which.[6] That is, in the model

(2) $$y_t = \gamma + aE_t y_{t+1} + cy_{t-1},$$

inclusion of the E_t operator before y_{t+1} indicates that the analyst has specified that a is the parameter that governs the magnitude of expectational influences of $E_t y_{t+1}$ on y_t while c is

[5] The importance and prevalence of such issues in monetary economics is stressed in McCallum (2003).
[6] It is possible, of course, to exclude one influence or the other by having either a or c equal to zero.

the parameter governing inertial effects of y_{t-1} on y_t.[7] It is crucial to recognize that in this framework agents are depicted as looking into the future while taking proper account of both expectational and inertial effects.[8] We could add exogenous variables, stochastic or deterministic, to this equation but doing so would have no major influence on the argument. The only state variable is y_{t-1} (which is predetermined) so the fundamental (sunspot-free) forward-looking linear solution is of the form

(3) $$y_t = \phi_0 + \phi_1 y_{t-1}.$$

Accordingly, $E_t y_{t+1} = \phi_0 + \phi_1(\phi_0 + \phi_1 y_{t-1})$ and simple undetermined-coefficient reasoning indicates that ϕ_0 and ϕ_1 must satisfy

(4) $$\phi_0 = \gamma + a\phi_0 + a\phi_1\phi_0$$

(5) $$\phi_1 = a\phi_1^2 + c.$$

For a given value of ϕ_1, (4) determines ϕ_0 uniquely but, clearly, (5) is satisfied by two values, which are

(6a) $$\phi_1^{(-)} = \frac{1-\sqrt{1-4ac}}{2a}$$

(6b) $$\phi_1^{(+)} = \frac{1+\sqrt{1-4ac}}{2a}.$$

We now ask, is there any connection between these two solutions and the correct identification of expectational and inertial components? Considering the special case in which c = 0, so the inertial component is absent, we see that the answer is arguably "yes."

---

[7] Then, a fully inertial specification with no significant role for expectations, i.e., a = 0, has as its solution $y_t = \gamma + cy_{t-1}$, not $y_t = (1/a)(y_{t-1} - cy_{t-2})$.

[8] It reflects a solution that Blanchard (1979, p.115) describes as "… a weighted average of two special solutions, a *backward solution* … and a *forward solution….*" Our argument is that careful attention to the model's causality properties results in a unique determination of the weights assigned to these special solutions. The lag-operator approach of Sargent (1979) is similar to, but more complete than, Blanchard's.

For in that case, $\phi_1^{(-)} = 0$ whereas $\phi_1^{(+)} = 1/a$. Thus the solution involving (6a) is appropriate whereas (6b) would suggest that causation is from y_{t-1} to y_t rather than from $E_t y_{t+1}$ to y_t.

That position is not accepted, however, by numerous analysts who take the position that expectations may depend on additional information variables, ones not included in the set of state variables implied by the model's specification. If, for example, y_{t-1} is such a variable and is included even when c = 0, then the conclusion that ϕ_1 should equal zero when c = 0 will not be accepted. More generally, this position argues for the eligibility as state variables "anything that agents decide to base their expectations on," including "sunspot variables," ones unrelated to the model at hand.

Let us instead consider, therefore, the situation in which a = 0, i.e., in which the importance of expectations in model (2) is nil. That is, we consider the contrasting special case in which c is non-zero but a equals zero. Then from (6b) we see that as $a \to 0$, we have $\phi_1^{(+)} \to \pm\infty$. By contrast, l'Hospital's rule shows that $\phi_1^{(-)}$ approaches c.[9] Thus we find that for this special case, as well as the one with c = 0, $\phi_1^{(-)}$ provides the a-priori correct value while $\phi_1^{(+)}$ implies a value that is incorrect in the sense of departing from the causality specification that has been built into the model. Furthermore, the two expressions for ϕ_1 are continuous functions of the basic parameters a and c of the model's structural relations. Thus for values of either c or a close to zero, the dynamic properties of the system, as determined by the value of ϕ_1, will be close to those known to be

[9] Both numerator and denominator of (6a) approach zero as $a \to 0$, but $d[1-\sqrt{1-4ac}]/da \to 2c$ while $d[2a]/da = 2$ so the expression in (6a) has a limiting value of c.

relevant if $\phi_1^{(-)}$ is adopted but not if $\phi_1^{(+)}$ were chosen for ϕ_1.[10] Accordingly, if we adopt the principle that the model's solution implies response functions that are continuous in the basic parameter values, we are justified in concluding that $\phi_1^{(-)}$ provides the answer in general, i.e., it identifies the solution that correctly represents the causal structure implied by model (3), in which both expectational and inertial components are potentially present.

In light of that contention, it is perhaps natural to ask, "to what causal dynamic specification does the solution $\phi_1^{(+)}$ pertain?" The answer is reasonably straightforward. Suppose the analyst ignores the expectation operator E_t in (2) and interprets the equation as a purely inertial model, writing it (with $\gamma = 0$) as

(2′) $$y_t = (1/a)y_{t-1} - (c/a)y_{t-2}.$$

Then for the special case with c = 0 he finds $y_t = (1/a)y_{t-1}$ as his solution, which is what he would find if he were using $\phi_1^{(+)}$. Generalizing, he might with $c \neq 0$ seek a solution to (2′) that features only one state variable, i.e.,

(3′) $$y_t = \psi y_{t-1}.$$

The latter would imply also that $y_t = \psi^2 y_{t-2}$, so substitution into (2′) would imply that ψ must satisfy $\psi^2 = (1/a)\psi - (c/a)$, which with $a \neq 0$ has the same form as (5). Accordingly, we see that (2′) is a second interpretation that gives rise to the quadratic equation (5). In fact, (2′) is the relation that is associated with the root $\phi_1^{(+)}$, not only in the special case c = 0, but also for cases with c close to zero, with a close to 0, and indeed

[10] This can be verified by numerical examples. Note for reference below that selection of the solution $\phi_1^{(-)}$ simultaneously implies that $\phi_1 \to 0$ as $c \to 0$ and that $\phi_1 \to c$ as $a \to 0$, whereas the other solution has $\phi_1^{(+)} \to 1/a$ as $c \to 0$ and $\phi_1^{(+)} \to \infty$ as $a \to 0$.

in general. Clearly, the causal structure is entirely different from that implied by the root $\phi_1^{(-)}$. The root $\phi_1^{(+)}$ pertains to an interpretation of model (2) as if it were model (2′), which has a different dynamic specification.

In sum, the solution based on expression (6b) for ϕ_1 is the solution to a model in which agents at time t make decisions about y_t on the basis of past values of y_{t-1} and y_{t-2}, plus the constraint implied by (3′), with expectations of y_{t+1} playing no role.[11] This is quite different from the model specified in (2), which depicts agents as choosing y_t values partly on the basis of y_{t-1} (inertial influences) and also (potentially) on expectations regarding y_{t+1}. From a structural point of view, these are two drastically different models. But once the analyst has decided which of the two models he is proposing, there is no ambiguity about its solution.[12]

Before moving on, we note that in the simple univariate model at hand the values $\phi_1^{(-)}$ and $\phi_1^{(+)}$ equal the eigenvalues of the dynamic system written in first-order form as well as possible values of the solution coefficient ϕ_1 in (6). The analogous equivalence does not prevail in multivariate models, but in the latter the system's eigenvalues continue to govern and describe the model's dynamic stability properties. And it is well known that in a multivariate system eigenvalues are continuous functions of the basic parameters of the model's structural equations.[13] Accordingly, it will be possible to relate different causal specifications of such a model to different groupings of system

[11] This is not a very sensible model, but it is what is implied by the solution using (6b). That is not at all a weakness for our argument, which by contrast posits (6a) as the appropriate expression.
[12] This *solution* might still fail to provide an *equilibrium* because of the failure of some transversality condition or some informational feasibility condition such as least-squares learnability (as developed by extensively by Evans and Honkapohja (2001)).
[13] See, e.g., Horn and Johnson (1985, pp. 539-540).

eigenvalues—groupings that imply different solutions for the multivariate counterpart of ϕ_1.

Indeed, it will be possible to generalize the conclusion obtained above—namely, that the appropriate RE solution can be identified as the one that results in a value of zero for the solution parameter ϕ_1 when the structural parameter c equals zero—by means of the multivariate counterpart of equation (2). It might be noted parenthetically that, in the univariate example above, this rule happens to coincide with adoption of the ϕ_1 value that is the smaller in absolute value of the two that satisfy the quadratic (5). Such a coincidence does not always obtain, however, in systems with more endogenous variables; and in such cases the appropriate solution must be based on the procedure just described. The mechanics of this procedure will be considered and demonstrated in Sections 4 and 5 below.

An example of model (2) can be provided by the classic analysis of capital accumulation. A simple version can be written as follows:

$$k_t + c_t = Ak_{t-1}^{\alpha} + (1-\delta)k_{t-1} \tag{7}$$

$$c_t^{-\sigma} = \beta c_{t+1}^{-\sigma}[A\alpha k_t^{\alpha-1} + (1-\delta)] \tag{8}$$

Here (7) is the resource constraint (with inelastic labor supply and Cobb-Douglas production function) while (8) is the intertemporal Euler equation (with constant intertemporal elasticity of substitution in consumption). One can linearize these and obtain

$$\hat{k}_t = \alpha_1\hat{c}_t + \alpha_2\hat{k}_{t-1} \qquad \alpha_1 = -\frac{c}{k} < 0, \qquad \alpha_2 = \alpha Ak^{\alpha-1} + 1 - \delta \tag{7′}$$

$$\hat{c}_t = \beta_1 E_t\hat{c}_{t+1} + \beta_2\hat{k}_t \qquad \beta_1 = 1, \qquad \beta_2 = -\frac{1}{\sigma}[\beta A\alpha(\alpha-1)k^{\alpha-1}] > 0 \tag{8′}$$

where the hatted values represent fractional deviations from steady state. Then we solve

(7′) to get $\hat{c}_t = \frac{\hat{k}_t - \alpha_2 \hat{k}_{t-1}}{\alpha_1}$, substitute into (8′), and rearrange, getting

(9) $$\hat{k}_t = \frac{\beta_1}{[1+\beta_1\alpha_2 - \alpha_1\beta_2]} E_t \hat{k}_{t+1} + \frac{\alpha_2}{[1+\beta_1\alpha_2 - \alpha_1\beta_2]} \hat{k}_{t-1}$$

which can be written as

(10) $$\hat{k}_t = A E_t \hat{k}_{t+1} + C\hat{k}_{t-1}.$$

The solution will be of form

(11) $$\hat{k}_t = \phi_1 \hat{k}_{t-1}$$

and we can consider the limiting cases mentioned above. Doing so, we find that for

$\alpha_2 \to 0$, $\phi_1^{(-)} \to 0$ but $\phi_1^{(+)} \to \frac{1}{1-\alpha_1\beta_2} \approx 1$. In addition, for $\beta_1 \to 0$ we have

$\phi_1^{(-)} \to C = \frac{\alpha_2}{1-\alpha_1\beta_2}$ whereas $\phi_1^{(+)} \to \frac{2}{0}$. These results are analogous to those on pp. 5-6

and again indicate clearly that $\phi_1^{(-)}$ is the solution that features continuity.

3. Sunspot Solutions

Before moving to multivariate models, let us consider "sunspot" solutions for model (2). These can be obtained by looking for solutions not of form (3) but more generally of form

(12) $$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 \xi_t$$

where ξ_t is any stationary stochastic process that has the property $E_{t-1}\xi_t = 0$.[14] To avoid

[14] The multivariate version of (12), considered in Section 4, corresponds to the complete set of solutions considered in Sims (2002) and Lubik and Schorfheide (2003). The latter authors apparently see it as desirable that Sims's method "… does not require the researcher to separate the list of endogenous variables … into 'jump' and 'predetermined' variables" (2003, p. 276). The position of the present paper is

unnecessary symbols, let us take $\phi_0 = 0$. Then we have

(13) $$E_t y_{t+1} = \phi_1(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 \xi_t) + \phi_2 y_{t-1} + 0.$$

Substituting these two expressions into model (2) then gives

(14) $$\phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 \xi_t = a[\phi_1(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 \xi_t) + \phi_2 y_{t-1}] + c y_{t-1}.$$

Accordingly, we have the undetermined-coefficient requirements

(15a) $$\phi_1 = a\phi_1^2 + a\phi_2 + c$$

(15b) $$\phi_2 = a\phi_1 \phi_2$$

(15c) $$\phi_3 = a\phi_1 \phi_3.$$

Now, the last two of these require that either $\phi_1 = 1/a$ or that $\phi_2 = \phi_3 = 0$. But in the latter case we have the same solutions as in Section 2. In the former case, ϕ_3 can be any number but then (15a) reduces to

(16) $$\frac{1}{a} = \frac{1}{a} + a\phi_2 + c,$$

that is, to $\phi_2 = -c/a$, which is not contradicted by (15b). So there is a sunspot solution

(17) $$y_t = \frac{1}{a} y_{t-1} - \frac{c}{a} y_{t-2} + \phi_3 \xi_t$$

for any value of ϕ_3. For a suitable range of values of a and c, each of these solutions will be dynamically stable. But whether it is stable or unstable, a solution of form (12) is essentially a stochastic extension of (2′) and thus reflects the same direction of causality as (3) with $\phi_1^{(+)}$ from (6b) above, which we have seen to be inconsistent with the dynamic specification of the model (2). Thus sunspot expressions of the form implied by (12)

---

basically just the opposite, assuming the desirability of having a structural model (which for Sims would perhaps not be the case).

with $E_{t-1}\xi_t = 0$ are not candidates for equilibria for the model (2), given that a and c are its parameters pertaining to expectational and inertial influences, respectively. Equivalently, no sunspot expressions of the indicated form yield candidate equilibria for model (2) with the specification that y_{t-1}, but not y_t, is predetermined. Under that specification, $a\phi_1 \neq 1$, $\phi_2 = \phi_3 = 0$, and the relevant candidate solution is (3) with (6a).

Given the logic of this argument, it seems very likely that our results can be extended to multivariate formulations of model (2). That is because the argument does not rely—as does analysis pertaining to E-stability and learnability as in Evans and Honkapohja (2001)—on quantitative values of system eigenvalues, which have different properties with respect to the crucial magnitude of their real parts for cases in which the number of endogenous variables exceeds 1. (On this point, see Horn and Johnson (1991, pp. 123, 130).) Instead, our argument relies only on the continuity of eigenvalues with respect to the model's parameters, given that we are able to identify each causality pattern with a single specification regarding predetermined variables, which in turn singles out a particular model solution. Nevertheless, it would seem desirable to consider multivariate formulations explicitly, to verify that the foregoing suggestion is in fact correct. Accordingly, we will sketch such an extension in the next section.

4. General Multivariate Formulation

To extend the results of the previous sections to a multivariate setting we will, following McCallum (2007), work with the following class of linear models:

(18) $y_t = A\, E_t y_{t+1} + C\, y_{t-1} + D\, u_t,$

where y_t is a m×1 vector of endogenous variables, A and C are m×m matrices of real numbers, D is m×n, and u_t is a n×1 vector of exogenous variables generated by a dynamically stable process

(19) $u_t = P\, u_{t-1} + \varepsilon_t$,

with ε_t a white noise vector and P a matrix with all eigenvalues less than 1.0 in modulus. It will not be assumed that A is invertible. In this formulation the endogenous variables in y_t are jump variables whereas their lagged values in y_{t-1} are predetermined, that is, dependent only on lagged values of exogenous or endogenous variables. This specification is useful for various reasons, the main one with respect to the issues at hand being that it is very broad and inclusive. In particular, any model satisfying the formulations of King and Watson (1998) or Klein (2001), can (with the use of auxiliary variables) be written in this form—and the form will accommodate any finite number of lags, expectational leads, and lags of expectational leads.[15] In that context, we consider solutions to model (18)-(19) of the form

(20) $y_t = \Omega\, y_{t-1} + \Gamma\, u_t$.

in which Ω is required to be real.[16] Then we have that $E_t y_{t+1} = \Omega(\Omega y_{t-1} + \Gamma u_t) + \Gamma R u_t$ and straightforward undetermined-coefficient reasoning shows that Ω and Γ must satisfy

(21) $A\Omega^2 - \Omega + C = 0$

(22) $\Gamma = A\Omega\Gamma + A\Gamma R + D$.

For any given Ω, (22) yields a unique Γ generically,[17] but there are many matrices of order $m\times m$ that solve (21) for Ω. Accordingly, the following analysis centers around equation (21).

---

[15] See McCallum (2007, p. 1379).

[16] A constant term can be defined by the coefficient on an exogenous variable that is a driftless random walk with innovation variance of zero.

[17] Generically, $I - R'\otimes[(I - A\Omega)^{-1}A]$ will be invertible, permitting solution for vec(Γ) using the identity vec(ABC) = $[C'\otimes A]$vec(B) that holds for any conformable A, B, C.

If A were invertible, we could express (21) in the first-order form

$$(23)\quad \begin{bmatrix} \Omega^2 \\ \Omega \end{bmatrix} = \begin{bmatrix} A^{-1} & -A^{-1}C \\ I & 0 \end{bmatrix} \begin{bmatrix} \Omega \\ I \end{bmatrix},$$

and proceed as in the well-known analysis of Blanchard and Khan (1980), which is based on the eigenvalues of the square matrix on the right-hand side of (23). With A singular, however, we proceed as follows. In place of (23), we write

$$(24)\quad \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \Omega^2 \\ \Omega \end{bmatrix} = \begin{bmatrix} I & -C \\ I & 0 \end{bmatrix} \begin{bmatrix} \Omega \\ I \end{bmatrix},$$

in which the first row reproduces the matrix quadratic (21). Let the 2m×2m matrices on the left and right sides of (19) be denoted $\bar{A}$ and $\bar{C}$, respectively. Then instead of focusing on the eigenvalues of $\bar{A}^{-1}\bar{C}$, which does not exist when A is singular, we instead solve for the (generalized) eigenvalues of the matrix pencil $[\bar{C} - \lambda\bar{A}]$, alternatively termed the (generalized) eigenvalues of $\bar{C}$ with respect to $\bar{A}$ (e.g., Uhlig (1999)). Thus instead of diagonalizing $\bar{A}^{-1}\bar{C}$, we use the Schur generalized decomposition, which serves the same purpose. Specifically, the Schur generalized decomposition theorem establishes that there exist unitary matrices Q and Z of order 2m×2m such that $Q\bar{C}Z = T$ and $Q\bar{A}Z = S$ with T and S triangular.[18] Then eigenvalues of the matrix pencil $[\bar{C} - \lambda\bar{A}]$ are defined as t_{ii}/s_{ii}. Some of these eigenvalues may be "infinite," in the sense that some s_{ii} may equal zero. This will be the case, indeed, whenever A and therefore $\bar{A}$ are of less than full rank since then S is also singular. All of the foregoing is true for any ordering of the eigenvalues and associated columns of Z

[18] Provided only that there exists some λ for which $\det[\bar{C} - \lambda\bar{A}] \neq 0$. See Klein (2000) or Golub and Van Loan (1996).

(and rows of Q). For the moment, let us temporarily focus on the arrangement that places the t_{ii}/s_{ii} in order of decreasing modulus, which will be referred to as the MOD ordering.

To begin the analysis, premultiply (24) by Q. Since $Q\bar{A} = SH$ and $Q\bar{C} = TH$, where $H \equiv Z^{-1}$, the resulting equation can be written as

$$(25) \quad \begin{bmatrix} S_{11} & 0 \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} \Omega^2 \\ \Omega \end{bmatrix} = \begin{bmatrix} T_{11} & 0 \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} \Omega \\ I \end{bmatrix}.$$

The first row of (25) reduces to $S_{11}(H_{11}\Omega + H_{12})\Omega = T_{11}(H_{11}\Omega + H_{12})$, so if H_{11} is invertible the latter can be used to solve for Ω, which is m×m, as

$$(26) \quad \Omega = -H_{11}^{-1} H_{12} = -H_{11}^{-1}(-H_{11}Z_{12}Z_{22}^{-1}) = Z_{12}Z_{22}^{-1},$$

where the second equality comes from the upper right-hand submatrix of the identity $HZ = I$, provided that H_{11} is invertible, which we assume without significant loss of generality.

As mentioned above, there are many solutions Ω to (21). These correspond to the $(2m)!/(m!)^2$ different combinations of the 2m eigenvalues taken m at a time, which result in different groupings of the 2m system eigenvalues into those that do and those that do not equal the eigenvalues of Ω. McCallum (2007) shows that for each such grouping the <u>other</u> system eigenvalues will equal the eigenvalues of the matrix F^{-1} where $F = (I - A\Omega)^{-1}A$ for the particular Ω at hand. Each grouping then amounts to a single specification as to which m of the system's 2m variables are predetermined, and therefore to a full specification of causality relationships among variables. Thus, when a particular causality specification is adopted, then a particular grouping is implied.

What grouping gives the solution such that $\Omega \to C$ as $A \to O$? In analogy to the univariate case, it is the same as the solution in which $C \to O$ implies $\Omega \to O$. In that case, and only in that case, all of the eigenvalues of Ω will equal zero when C = O. Thus the only solution that reflects the modeller's causality/predetermination assumptions is the one for which $\Omega \to O$ as elements of $C \to 0$. It can be identified by replacing C by κC in all equations and then letting κ decrease from 1 to 0. A plot or table of the eigenvalues for various κ will indicate which solution—i.e., which Ω—has this property. This procedure, mentioned previously in McCallum (2004), will be illustrated below in Section 5.

Given the foregoing, sunspot solutions can be introduced for model (18)(19) by looking for solutions of form

(27) $$y_t = \Omega y_{t-1} + \Phi_1 y_{t-2} + \Phi_2 \xi_t$$

where I have removed the exogenous vector u_t but added y_{t-2} and the m×1 sunspot vector ξ_t that has the property $E_{t-1}\xi_t = 0$. Then the analysis proceeds in the same manner as in Section 3, showing that either $A\Omega = I$ or $\Phi_1 = \Phi_2 = O$. Suppose $A\Omega \neq I$. Then it is required that $\Phi_1 = O$, $\Phi_2 = O$, and we have the solution given in Section 4 with no sunspot vector. But if $A\Omega = I$, $\Phi_1 \neq O$ and Φ_2 is arbitrary. These solutions—there are many Ωs that satisfy $A\Omega = I$—include a sunspot term, but each of them differs from the one that has $\Omega \to O$ as $C \to O$, and the latter is the solution implied by the model. So adoption of a full causality specification rules out sunspots of the form implied by (27).

5. Two-by-Two Examples

Here the purpose is to examine a pair of examples in which there are two endogenous variables, y_{1t} and y_{2t}. For the first we use formulation (18) with the following parameter specification:

(28) $A = \begin{bmatrix} 0.3 & 0.01 \\ 0.0 & 0.60 \end{bmatrix} \qquad C = \begin{bmatrix} 0.1 & 0.0 \\ 0.0 & 0.2 \end{bmatrix}.$

Here the example has two sectors that are related only by the appearance of $E_t y_{2t+1}$ in the equation for y_{1t} with a small coefficient a_{12} set equal to 0.01. If that coefficient were zero, then the system would be a pair of univariate models. Their system eigenvalues would be 3.2301, 0.1032 for the first and 1.4343, 0.2324 for the second. Thus the MOD solutions would be $y_{1t} = 0.1032y_{1t-1}$ and $y_{2t} = 0.2324y_{2t-1}$, both of which are dynamically stable. With the slight interaction provided by $a_{12} = 0.01$, the bivariate system has almost exactly the same properties, the solution being of form (15) with $\Omega = \begin{bmatrix} 0.1032 & 0.0006 \\ 0.0000 & 0.2324 \end{bmatrix}$.

Suppose, however, that we consider one of the other solutions for Ω. Computationally, we can do this by using the code QZSWITCH to reverse (relative to the MOD ordering) the positions of the m-th and (m+1)-th eigenvalues, a step that has the effect of switching positions of two eigenvectors that are columns of the Z matrix that appears in equation (22) above. (More extensive reorderings can be effected by repeated application of QZSWITCH, suitably directed.) If this is done for the system at hand, with 0.2324 placed ahead of 1.4343 in the ordering that was previously MOD, the resulting Ω matrix becomes $\Omega = \begin{bmatrix} 0.1032 & 0.0382 \\ 0.0000 & 1.4343 \end{bmatrix}$. What is one to make of this, i.e., what is the implied behavior? Here again the y_{2t} process is autonomous, but now the solution for y_{2t} is explosive and entirely "backward looking," as with the solution (6b) in the example given at the start of Section 2. This occurs because the list of relevant predetermined variables for the system has become y_{1t-1} and y_{2t}, rather than y_{1t-1} and y_{2t-1}. In the resulting solution, the variable y_{1t} is dependent upon $E_t y_{2t+1}$ via ω_{12} so it fluctuates in a stable manner about the explosive solution for y_{2t}.[19]

In a bivariate system of form (18) for variables y_{1t} and y_{2t}, the six possible pairs of predetermined variables are (y_{1t-1}, y_{2t-1}), (y_{1t-1}, y_{2t}), (y_{1t-1}, y_{1t}), (y_{1t}, y_{2t-1}), (y_{2t}, y_{2t-1}), and (y_{1t}, y_{2t}). The first of these is the model-implied specification and the sixth one pertains to the bivariate formulation in which both structural equations are interpreted as backward looking, as in the univariate example presented above in (2′). It appears that the second and fourth can also be given intelligible interpretations, but possibly not the third and fifth. In any case, it is the second that is picked out by the rearrangement that is reported above.

A tentative conclusion, then, is that the interpretation of alternative solutions, obtained by alternative eigenvalue orderings, is in some cases clearcut but in others may be somewhat problematical. But in a certain class of cases, an eigenvalue reordering is exactly what is needed to make sense of a model's causality specification. The relevant class of cases is that in which it is not true that, with the MOD ordering, $\Omega \rightarrow O$ as all elements of C approach zeros—which condition we have argued above is necessary for the MOD solution to reflect the modeller's own causality/predetermination assumptions. To illustrate the point, let us consider a second example, this one from McCallum (2004).

The last-mentioned paper illustrates that a determinate solution—in the sense of being the only dynamically stable solution—may differ from the MSV (minimum state variable) solution, as defined by McCallum (1983, 2004), for the model at hand.[20] Since

[19] Here ω_{ij} denotes the i,j element of Ω.

[20] It must be noted that there are two different terminological conventions pertaining to the MSV concept. Ours makes the MSV concept uniquely equal to the one for which $\Omega \rightarrow O$ as $C \rightarrow O$. A more general concept is used by Evans and Honkapohja (2001, p. 194), who are careful to note the distinction.

to be determinate a solution must have the MOD ordering, this example illustrates that MSV and MOD solutions can differ. This is a rare but possible occurrence. The first example in the paper is given by equation (13) with the following A, C, and D matrices:[21]

$$(29) \qquad A = \begin{bmatrix} -1.5 & 1.2 \\ 0.5 & -1.3 \end{bmatrix} \qquad C = \begin{bmatrix} 1.2 & 0.5 \\ 0.3 & 1.6 \end{bmatrix} \qquad D = \begin{bmatrix} 0.0 & 0.0 \\ 0.0 & 0.0 \end{bmatrix}.$$

The system eigenvalues are −2.7022, 1.0887, −0.9365, and 0.4759, so there is a single stable solution. This solution, however, is not the one for which the matrix Ω approaches a null matrix as C or κC approaches zero. That this is the case is shown in Figure 1, where the moduli of the four eigenvalues are plotted against κ.[22] At a value of κ close to 0.7 the paths of the second-largest and third-largest moduli cross, so that it is the second and forth largest that approach zeros as κ approaches zero.[23] If one takes seriously the structural specification of the model, therefore, he needs to obtain the Ω matrix that is implied (with the value κ = 1) for the second- and fourth-largest eigenvalues.

How can that solution be obtained computationally? The answer is provided above; it is to modify the RE solution code by means of the QZSWITCH routine, thereby switching positions for the second and third eigenvalues. Doing so gives the matrix:

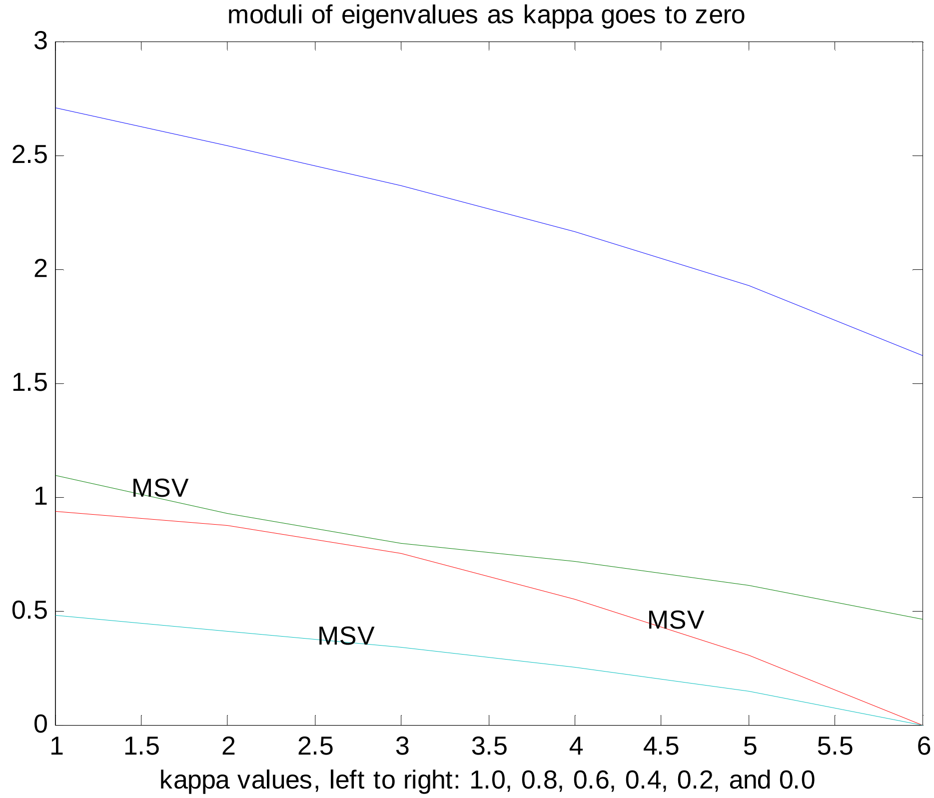
$$(30) \qquad \Omega = \begin{bmatrix} 0.6995 & 0.4489 \\ 0.1939 & 0.8651 \end{bmatrix}.$$

The eigenvalues of this matrix are, of course, 1.0887 and 0.4759. Accordingly, the system is dynamically explosive. This should not be taken as a flaw; it is an indication that the specified model, when care is taken with its dynamic causality structure, implies

---

[21] Please note that $c_{21} = 0.3$, not the value 0.5 that is incorrectly given in the published paper (2004, p. 58).
[22] Note that the index plotted on the horizontal axis of Fig. 1 equals 1−5(κ−1).
[23] These eigenvalue paths are labeled "MSV" in an attempt to emphasize the crossing, which the MATLAB plotter does not recognize. There is no danger of ambiguity about the fact of this crossing, since one of the eigenvalues is negative and the other positive. See McCallum (2004, p. 58).

Figure 1



explosive behavior. If the model's specification implies that there is a transversality condition that is violated by this explosion, then it is also implied that the solution does not represent a dynamic equilibrium. It represents a situation that should be avoided, presumably, if there is a policy component of the model that could be used in that way.

6. Conclusion

Let us conclude with a brief statement, in words, of what the paper has argued. Consider a rational expectations (RE) model that includes a relationship between the variables x_t and z_{t+1}, possibly the same variable at different dates. For such a model to be useful as a guide to actual behavior, it must specify whether x_t is influenced (via this relationship) by the expectation at t of z_{t+1} or, alternatively, that z_{t+1} is influenced (via some inertial mechanism) by x_t (i.e., that z_t is influenced by x_{t-1}). These are very different phenomena, between which an analyst concerned with structure will need to distinguish. (The distinctions will be provided automatically if the model is based on optimizing analysis of agents' choice problems.) The present paper shows that, for a very broad class of multivariate linear RE models, distinct causal specifications of these two types will be uniquely associated with distinct RE solutions. Furthermore, the different solutions will result operationally from different specifications concerning which of the model's variables are predetermined state variables. It follows that, for a given structural specification and with a natural continuity assumption, there is a single implied solution that is a candidate for a refinement equilibrium for the model at hand. Furthermore, this particular solution does not involve "sunspot" phenomena. Thus the paper's message clashes strongly with the conclusions of a huge literature on sunspots, indeterminacy, and related phenomena that can appear in RE models. It does not clash, by contrast, with the notion that, for a RE solution to be plausible, it must not violate any relevant transversality condition and/or must be least-squares learnable.

References

Blanchard, O.J. "Backward and Forward Solutions for Economies with Rational Expectations," American Economic Review 69, May 1979, 114-118.

Blanchard, O.J., and C.M. Kahn."The Solution of Linear Difference Models Under Rational Expectations," Econometrica 48, 1980, 1305-1311.

Calvo, G. "On the Indeterminacy of Interest Rates and Wages with Perfect Foresight," Journal of Economic Theory 19, 1978, 321-337.

Cho, S., and B.T. McCallum. "Another Weakness of 'Determinacy' as a Selection Criterion for Rational Expectations Models," Economics Letters 104, 2009, 17-19.

Cochrane, J.H. "Inflation Determination with Taylor Rules: A Critical Review." NBER Working Paper 13409, 2007.

Evans, G.W., and S. Honkapohja. Learning and Expectations in Macroeconomics. Princeton Univ. Press, 2001.

Golub, G.H., and C.F. VanLoan. Matrix Computations, 3rd ed. Johns Hopkins University Press, 1996.

Horn, R.A., and C. R. Johnson. Matrix Analysis. Cambridge University Press, 1985.

__________________. Topics in Matrix Analysis. Cambridge University Press, 1991.

King, R.G., and M.W. Watson. "The Solution of Singular Linear Difference Systems Under Rational Expectations," International Economic Review 39, 1998, 1015-1026.

Klein, P. “Using the Generalized Schur Form to Solve a Multivariate Linear Rational Expectations Model,” Journal of Economic Dynamics and Control 24, 2000, 1405-1423.

Lubik, Thomas A., and Frank Schorfheide. “Computing Sunspot Equilibria in Linear Rational Expectations Models,” Journal of Economic Dynamics and Control 28, 2003, 273-285.

McCallum, B.T. “On Nonuniqueness in Linear Rational Expectations Models: An Attempt at Perspective,” Journal of Monetary Economics 11, 1983, 139-168.

_____________, “Multiple-Solution Indeterminacies in Monetary Policy Analysis,” Journal of Monetary Economics 50, 2003, 1153-1175.

______________“On the Relationship Between Determinate and MSV Solutions in Linear RE Models,” Economics Letters 84, 2004, 55-60.

_____________. “E-stability vis-a-vis Determinacy Results for a Broad Class of Linear Rational Expectations Models,” Journal of Economic Dynamics and Control 31, 2007, 1376-1391.

Sargent, T.J. Macroeconomic Theory. Academic Press, 1979 (2nd edition 1987).

Simon, H.A. “Causal Ordering and Identifiability,” in Studies in Econometric Method, Cowles Commission Monograph No. 14, eds. W.C. Hood and T.C. Koopmans. John Wiley & Sons, 1953.

Sims, C.A. “Solving Linear rational Expectations Models,” Computational Economics 20, October 2002, 1-20.

Uhlig, H."A Toolkit for Analyzing Nonlinear Dynamic Stochastic Models Easily," in Computational Methods for the Study of Dynamic Economies, R. Marimon and A. Scott, eds. Oxford University Press, 1999.

Zellner, A. "Causality and Econometrics," Carnegie-Rochester Conference Series on Public Policy, Vol 10. North-Holland Pub. Co., 1979.