Topic 2: Solution Methods for Rational Expectations Models

Having described econometric methods for measuring the shocks that hit the macroeconomy and their dynamic effects, we now turn to developing theoretical models that can explain these patterns. It should be clear that this task requires models with explicit dynamics and with stochastic shocks. Macroeconomic models tend to have both backward- and forward-looking dynamics. Backward-looking dynamics stem, for instance, from identities linking today's capital stock with last period's capital stock and this period's investment. Forward-looking dynamics generally require an assumption about how people formulate expectations.

Rational Expectations and Macroeconomics

Almost all economic transactions rely crucially on the fact that the economy is not a "one-period game." In the language of macroeconomists, most economic decisions have an *intertemporal* element to them. So, a key issue in macroeconomic theory is how people formulate expectations of future economic variables in the presence of uncertainty. Prior to the 1970s, this aspect of macro theory was largely *ad hoc*. Different researchers took different approaches, but generally it was assumed that agents used some simple extrapolative rule whereby the expected future value of a variable was close to some weighted average of its recent past values. However, such models were widely criticised in the 1970s by economists such as Robert Lucas and Thomas Sargent. Lucas and Sargent instead promoted the use of an alternative approach which they called "rational expectations." This approach had been introduced in an important 1961 paper by John Muth, and this original paper is worth reading even if it doesn't have much discussion of the implications for macroeconomics.[1]

The idea that agents expectations are somehow "rational" has various possible interpretations. However, when economists say that agents in a model have rational expectations, they usually mean two very specific things:

- They use publicly available information in an efficient manner. Thus, they do not make systematic mistakes when formulating expectations.
- They understand the structure of the model economy and base their expectations of variables on this knowledge.

[1] John Muth. "Rational Expectations and the Theory of Price Movements," *Econometrica*, July 1961.

To many economists, this is a natural baseline assumption: We usually assume agents behave in an optimal fashion, so why would we assume that the agents don't understand the structure of the economy, and formulate expectations in some sub-optimal fashion. That said, rational expectations models generally produce quite strong predictions, and these can be tested. Ultimately, any assessment of a rational expectations model must be based upon its ability to fit the relevant macro data.

First-Order Stochastic Difference Equations

The simplest rational expectations model is the *first-order stochastic difference equation*, which takes the form

$$y_t = x_t + bE_t y_{t+1} \tag{1}$$

I know you have seen some examples of this type of equation already, so I will only briefly recap here on the general approach taken for solving this model. Its solution is derived using a technique called *repeated substitution*. This works as follows. Equation (1) holds in all periods, so under the assumption of rational expectations, the agents in the economy understand the equation and formulate their expectation in a way that is consistent with it:

$$E_t y_{t+1} = E_t x_{t+1} + bE_t y_{t+2} \tag{2}$$

Substituting this into the previous equation, we get

$$y_t = x_t + bE_t x_{t+1} + b^2 E_t y_{t+2} \tag{3}$$

Repeating this method by substituting in for $E_t y_{t+2}$, and then $E_t y_{t+3}$ and so on, we get a general solution of the form

$$y_t = x_t + bE_t x_{t+1} + b^2 E_t x_{t+2} + ... + b^{N-1} E_t x_{t+N-1} + b^N E_t y_{t+N} \tag{4}$$

which can be written in more compact form as

$$y_t = \sum_{k=0}^{N-1} b^k E_t x_{t+k} + b^N E_t y_{t+N} \tag{5}$$

Usually, it is assumed that

$$\lim_{N \to \infty} b^N E_t y_{t+N} = 0 \tag{6}$$

So the solution is

$$y_t = \sum_{k=0}^{\infty} b^k E_t x_{t+k} \tag{7}$$

This solution underlies the logic of a very large amount of modern macroeconomics.

"Backward" Solutions

The model

$$y_t = x_t + bE_t y_{t+1} \tag{8}$$

can also be written as

$$y_t = x_t + by_{t+1} + b\epsilon_{t+1} \tag{9}$$

where ϵ_{t+1} is a forecast error that cannot be predicted at date t. Moving the time subscripts back one period and re-arranging this becomes

$$y_t = b^{-1}y_{t-1} - b^{-1}x_{t-1} - \epsilon_t \tag{10}$$

This backward-looking equation which can also be solved via repeated substitution to give

$$y_t = -\sum_{k=0}^{\infty} b^{-k}\epsilon_{t-k} - \sum_{k=1}^{\infty} b^{-k}x_{t-k} \tag{11}$$

Equations (7) and (11) are both correct solutions to the first-order stochastic difference equation. However, which one we choose to emphasize will depend on the value of the parameter b. If $|b| > 1$, then the weights on future values of x_t in equation (7) will explode. In this case, it is most likely that equation (6) will not hold so the forward solution does not converge to a finite sum. Even if it does, the idea that today's value of y_t depends more on values of x_t far in the distant future than it does on today's values is not one that we would be comfortable with. So, in this case, equation (11) is the solution to focus on for practical applications. In contrast if $|b| < 1$ then the weights in (11) are explosive and equation (7) is the one to focus on.

From Structural to Reduced Form Relationships

Equation (7) provides useful insights into how the variable y_t is determined. However, without some assumptions about how x_t evolves over time, it cannot be used to give precise predictions about they dynamics of y_t. Ideally, we want to be able to simulate the behaviour of y_t on the computer. One reason there is a strong linkage between DSGE modelling and VARs is that this question is usually addressed by assuming that the exogenous "driving variables" such as x_t are generated by backward-looking time series models like VARs.

Consider for instance the case where the process driving x_t is

$$x_t = \rho x_{t-1} + \epsilon_t \tag{12}$$

where $|\rho| < 1$. In this case, we have

$$E_t x_{t+k} = \rho^k x_t \tag{13}$$

Now the model's solution can be written as

$$y_t = \left[\sum_{k=0}^{\infty} (b\rho)^k \right] x_t \tag{14}$$

Because $|b\rho| < 1$, the infinite sum converges to

$$\sum_{k=0}^{\infty} (b\rho)^k = \frac{1}{1 - b\rho} \tag{15}$$

Remember this is the identity that generates the famous Keynesian multiplier formula. So, in this case, the model solution is

$$y_t = \frac{1}{1 - b\rho} x_t \tag{16}$$

Macroeconomists call this a *reduced-form* solution for the model: Together with equation (12), it can easily be simulated using a computer program to generate artificial time series for the behavior of y_t and x_t which can then be compared with real data.

While this example is obviously a relatively simple one, it illustrates the general principal for getting predictions from DSGE models:

- Obtain *structural* solutions involving expectations of future driving variables (in this case, equation 7).
- Make assumptions about the time series process for the *driving variables* (in this case equation 12)
- Solve for a *reduced-form* solution than can be simulated on the computer along with the driving variables (in this case equation 16)

Before moving on, note that the reduced-form of this model also has a VAR-like representation, which can be shown as follows:

$$y_t = \frac{1}{1 - b\rho} (\rho x_{t-1} + \epsilon_t) \tag{17}$$

$$= \rho y_{t-1} + \frac{1}{1 - b\rho} \epsilon_t \tag{18}$$

So both the x_t and y_t series have purely backward-looking representations. Even this simple model helps to explain how theoretical models tend to predict that the data can be described well using a VAR.

The Lucas Critique

Reduced-form representations like (16) are necessary to simulate the time series implied by the structural model (7). However, it is crucial to keep in mind that equation (7) is *always* true for this model, while the reduced-form representation (16) depends on the process for x_t taking a particular form. Should that process change, the reduced-form process will change.

This idea—that coefficients of reduced-form relationships depend crucially on how agents formulate expectations about the future—is one of the most important themes in modern macroeconomics. In particular, in a famous paper, rational expectations pioneer Robert Lucas pointed out that the assumption of rational expectations implied that these coefficients would change if expectations about the future changed.[2] Lucas stressed that this could make reduced-form econometric models based on historical data useless for policy analysis. This problem is now known as the *Lucas critique* of econometric models.

Consider, for example, a simple "permanent income" model in which consumption depends on a present discounted value of after-tax income

$$c_t = \gamma \sum_{k=0}^{\infty} \beta^k E_t y_{t+k} \tag{19}$$

and suppose that income has followed the process

$$y_t = (1+g)\, y_{t-1} + \epsilon_t \tag{20}$$

In this case, we have

$$E_t y_{t+k} = (1+g)^k \, y_t \tag{21}$$

So the reduced-form representation is

$$c_t = \gamma \left[\sum_{k=0}^{\infty} \left(\beta\,(1+g)\right)^k \right] y_t \tag{22}$$

Assuming that $\beta\,(1+g) < 1$, this becomes

$$c_t = \frac{\gamma}{1 - \beta\,(1+g)} y_t \tag{23}$$

Now suppose that the government is thinking of introducing a temporary income tax cut. As noted above, we can consider y_t to be after-tax labour income, so it would be

[2] Robert Lucas, "Econometric Policy Evaluation: A Critique," *Carnegie-Rochester Series on Public Policy*, Vol. 1, pages 19-46.

temporarily boosted by the tax cut. Now suppose the policy-maker wants an estimate of the likely effect on consumption of the tax cut. They may get their economic advisers to run a regression of consumption on after-tax income. If, in the past, consumers had generally expected income growth of g, then the econometric regressions will report a coefficient of approximately $\frac{\gamma}{1-\beta(1+g)}$ on income. So, the economic adviser might conclude that for each extra dollar of income produced by the tax cut, there will be an increase in consumption of $\frac{\gamma}{1-\beta(1+g)}$ dollars.

However, if households have rational expectations and operate according to the structural model (19) then the true effect of the tax cut could be a lot smaller. For instance, if the tax cut is only expected to boost this period's income, and to disappear tomorrow, then each dollar of tax cut will produce only γ dollars of extra consumption. The difference between the true effect and the economic advisor's supposedly "scientific" regression-based forecast could be substantial. Plugging in some reasonable numbers, suppose $\beta = 0.95$ and $g = 0.02$. In this case, the economic advisor concludes that the effect of a dollar of tax cuts is an extra 32γ $(=\frac{\gamma}{1-\beta(1+g)})$ dollars of consumption. In reality, the tax cut will produce only an extra γ dollars of extra consumption. Obviously, being off by a factor of 32 constitutes a big mistake in assessing the effect of this policy.

The original Lucas paper is relatively difficult and not available on the internet. Instead, I have put a paper on the reading list by Thomas Sargent which is short and readable and has a nice example of investment behaviour with temporary tax policies.[3]

The Lucas Critique and the Limitations of VAR Analysis

The tax cut example gets the logic of the critique across but perhaps not its generality. Today's DSGE models feature policy equations that describe how monetary policy is set via rules relating interest rates to inflation and unemployment; how fiscal variables depends on other macro variables; what the exchange rate regime is. These models all feature rational expectations, so changes to these policy rules will be expected to alter the reduced-form VAR-like structures associated with these economies.

This issue is an important "selling point" for modern DSGE models. These models can explain why VARs fit the data well, but they can be considered superior tools for policy analysis. They explain how reduced-form VAR-like equations are generated by the

[3]Thomas Sargent (1980). "Rational Expectations and the Reconstruction of Macroeconomics" *Federal Reserve Bank of Minneapolis Quarterly Review.*

processes underlying policy and other driving variables. However, while VAR models do not allow reduced-form correlations change over time, a fully specified DSGE model can explain such patterns as the result of structural changes in policy rules. In this sense, the DSGE approach is an attempt to produce models robust to the Lucas critique.

Systems of Rational Expectations Equations

So far, we have only looked at a single equation linking two variables. However, it turns out that the logic of the first-order stochastic difference equation underlies the solution methodology for just about all rational expectations models. Suppose one has a vector of variables

$$Z_t = \begin{pmatrix} z_{1t} \\ z_{2t} \\ . \\ z_{nt} \end{pmatrix} \tag{24}$$

It turns out that a lot of macroeconomic models can be represented by an equation of the form

$$Z_t = BE_t Z_{t+1} + X_t \tag{25}$$

where B is an $n \times n$ matrix. The logic of repeated substitution can also be applied to this model, to give a solution of the form

$$Z_t = \sum_{k=0}^{\infty} B^k E_t X_{t+k} \tag{26}$$

As with the single-equation model, this will only give a stable non-explosive solution under certain conditions. To understand these conditions, we need to talk about ....

Eigenvalues

A value λ_i is an eigenvalue of the matrix B if there exists a vector e_i (known as an eigenvector) such that

$$Be_i = \lambda_i e_i \tag{27}$$

Most $n \times n$ matrices have n distinct eigenvalues. Denote by P the matrix that has as its columns n eigenvectors corresponding to these eigenvalues. In this case,

$$BP = P\Omega \tag{28}$$

where

$$\Omega = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & & 0 \\ 0 & 0 & 0 & \lambda_n \end{pmatrix} \tag{29}$$

is a diagonal matrix of eigenvalues. Note now that this equation implies that

$$B = P\Omega P^{-1} \tag{30}$$

This tells us something about the relationship between eigenvalues and higher powers of B because

$$B^n = P\Omega^n P^{-1} = P \begin{pmatrix} \lambda_1^n & 0 & 0 & 0 \\ 0 & \lambda_2^n & 0 & 0 \\ 0 & 0 & & 0 \\ 0 & 0 & 0 & \lambda_n^n \end{pmatrix} P^{-1} \tag{31}$$

So, the difference between lower and higher powers of B is that the higher powers depend on the eigenvalues taken to the power of n. If all of the eigenvalues are inside the unit circle (i.e. less than one in absolute value) then all of the entries in B^n will tend towards zero as $n \to \infty$. So, a condition that ensures that a model of the form of equation (25) has a unique stable forward-looking solution is that the eigenvalues of B are all inside the unit circle.[4]

Generality of First-Order Matrix Formulation

Remember how the first-order matrix formulation of the VAR model could be used to represent models with more than one lag. At first glance, it looks as though equation (25) only allows for first-order purely forward-looking difference equations. However, it turns out that the same "companion matrix" trick can also be applied here, so that this formulation can apply to more general models. Consider, for instance, the model

$$y_t = ay_{t-1} + bE_t y_{t+1} + x_t \tag{32}$$

[4] To keep things manageable, I'm simplifying from the general case quite a bit here. Not all matrices have distinct eigenvalues, and this prevents them from being diagonalizable in the manner discussed here. But the eigenvalues still determine the stability of the system even in this case.

This model has a forward-looking and backward-looking element. To see that this model can still fit within the first-order matrix formulation, note that it can be re-written as

$$y_{t-1} = \frac{1}{a} y_t - \frac{b}{a} E_t y_{t+1} - \frac{1}{a} x_t \tag{33}$$

This can be expressed in first-order matrix form as

$$\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\frac{b}{a} & \frac{1}{a} \end{pmatrix} E_t \begin{pmatrix} y_{t+1} \\ y_t \end{pmatrix} + \begin{pmatrix} 0 \\ -\frac{1}{a} x_t \end{pmatrix} \tag{34}$$

General Solution a la Blanchard and Kahn (1980)

The idea of expressing rational expectations models in the form (25) was first discussed by Blanchard and Kahn (1980).[5] One point stressed in their paper is that we do not necessarily want all of the eigenvalues of the B matrix to be less than one. Consider the second-order difference equation model just discussed. In this case, the Z_t vector contains a term (y_{t-1}) for which a forward-looking solution is not appropriate: y_{t-1} is predetermined (to use the terminology of Blanchard and Kahn). This means it can't jump up and down with expectations determined at time t. For a model like this, we would expect B to have eigenvalues both inside and outside the unit circle.

To provide a more concrete illustration of this idea, re-write (25) as

$$Z_t = \left(P \Omega P^{-1}\right) E_t Z_{t+1} + X_t \tag{35}$$

Now multiply both sides by P^{-1} to get

$$P^{-1} Z_t = \Omega E_t \left(P^{-1} Z_{t+1}\right) + P^{-1} X_t \tag{36}$$

Defining new vectors of variables

$$W_t = P^{-1} Z_t \tag{37}$$

$$V_t = P^{-1} X_t \tag{38}$$

The model can be re-written as

$$W_t = \Omega E_t W_{t+1} + V_t \tag{39}$$

[5] I've put this paper on the reading list for completeness but am not recommending that you have to go through all its details.

This model is just n separate equations of the form

$$w_{it} = \lambda_i E_t w_{i,t+1} + v_{it} \tag{40}$$

Obviously, we can then solve each of these separate equations in the appropriate manner: Those with $|\lambda_i| < 1$ can be solved forward and those with $|\lambda_i| > 1$ can be solved backwards. One solutions for W_t are obtained, we can then obtain solutions for the variables of interest by calculating $Z_t = PW_t$.

The Binder-Pesaran Method

The above is more or less a standard description of the analytics of multi-equation rational expectations models. However, from a practical perspective, this approach has a few drawbacks as a way of formulating and solving these models:

- Re-writing models with forward- and backward-looking dynamics into the first-order Blanchard-Kahn form (e.g. going from equation 33 to equation 34) is a bit of pain
- Sorting through the model and solving subcomponents separately depending on their eigenvalues being inside or outside the unit circle, and then transforming solutions from W_t to Z_t is also tedious.

For these reasons, I'll describe a method that I prefer for solving these models. This method was introduced by Binder and Pesaran (1995). Their paper is tough going, but you might want to read the relevant parts of it if you want some more details.

The first advantage of this model is that it deals with these models in their more intuitive form, featuring forward- and backward-looking dynamics separately as

$$Z_t = AZ_{t-1} + BE_t Z_{t+1} + X_t \tag{41}$$

Again, the restriction to one-lag one-lead form is only apparent, and the companion matrix trick can be used to allow this model to represent models with n leads and lags. In this sense, this equation summarizes all possible linear rational expectations models. Binder and Pesaran's solution method is to find a matrix C such that $W_t = Z_t - CZ_{t-1}$ obeys a first-order matrix equation of the form

$$W_t = FE_t W_{t+1} + GX_t \tag{42}$$

In other words, it transforms solving the "second-order" system in equation (41) into a simpler first-order system.

What must the matrix C be? Using the fact that

$$Z_t = W_t + CZ_{t-1} \tag{43}$$

equation (41) can be re-written as

$$\begin{aligned} W_t + CZ_{t-1} &= AZ_{t-1} + B\left(E_t W_{t+1} + CZ_t\right) + X_t && (44)\\ &= AZ_{t-1} + B\left(E_t W_{t+1} + C\left(W_t + CZ_{t-1}\right)\right) + X_t && (45)\end{aligned}$$

This re-arranges to

$$\left(I - BC\right) W_t = BE_t W_{t+1} + \left(BC^2 - C + A\right) Z_{t-1} + X_t \tag{46}$$

Because the definition of C is that it is the matrix that such that W_t follows a first-order forward-looking matrix equation (with no extra Z_{t-1} terms) it follows that

$$BC^2 - C + A = 0 \tag{47}$$

This "matrix quadratic equation" can be solved to give C.[6] Once this is obtained, we have

$$W_t = FE_t W_{t+1} + GX_t \tag{48}$$

where

$$\begin{aligned} F &= \left(I - BC\right)^{-1} B && (49)\\ G &= \left(I - BC\right)^{-1} && (50)\end{aligned}$$

Assuming the all the eigenvalues of F are inside the unit circle, this has a forward-looking solution

$$W_t = \sum_{k=0}^{\infty} F^k E_t \left(GX_{t+k}\right) \tag{51}$$

which can be written in terms of the original equation as

$$Z_t = CZ_{t-1} + \sum_{k=0}^{\infty} F^k E_t \left(GX_{t+k}\right) \tag{52}$$

I like this method because it allow us to write models in a more intuitive form, and directly delivers the type of solution we want. Also, for well-specified models, the regularity conditions (eigenvalues of C and F being inside the unit circle) will generally hold, so one doesn't need to worry too much about eigenvalues.

[6] For instance, one can use the fact that $C = BC^2 + A$, to solve for it iteratively as follows. Provide an initial guess, say $C_0 = I$, and then iterate on $C_n = BC_{n-1}^2 + A$ until all the entries in C_n converge.

Reduced-Form Representation of DSGE Models

The Binder-Pesaran solution makes it very easy to derive reduced-form representations that can be simulated and then compared with real data. For instance, suppose we assume that the driving variables X_t follow a VAR representation of the form

$$X_t = DX_{t-1} + \epsilon_t \tag{53}$$

where D has eigenvalues inside the unit circle. The transformed driving variables GX_t also follow a VAR process of the form

$$GX_t = \left(GDG^{-1}\right)(GX_{t-1}) + G\epsilon_t = R\,(GX_{t-1}) + G\epsilon_t \tag{54}$$

This implies that

$$E_t GX_{t+k} = R^k GX_t \tag{55}$$

So the model has a solution of the form

$$Z_t = CZ_{t-1} + \left(\sum_{k=0}^{\infty} F^k R^k\right) GX_t \tag{56}$$

This infinite geometric sum of matrices looks a lot like the "multiplier-like" geometric sum from our earlier example and indeed it is. If the eigenvalues of FD are less than one (as they will be if both F and D have eigenvalues less then one themselves) then this infinite sum converges to

$$\sum_{k=0}^{\infty} F^k R^k = (I - FR)^{-1} \tag{57}$$

So, the model has a reduced-form representation

$$Z_t = CZ_{t-1} + (I - FR)^{-1} GX_t \tag{58}$$

which can be simulated along with the VAR process for the driving variables.

This provides a relatively simple recipe for simulating DSGE models. Specify the A and B matrices; solve for C and F; specify a VAR process for the driving variables; and then obtain the reduced-form representations. Binder and Pesaran have also provided GAUSS code that allows you to set up the model in the form

$$HZ_t = MZ_{t-1} + KE_t Z_{t+1} + X_t \tag{59}$$

$$X_t = LX_{t-1} + \epsilon_t \tag{60}$$

and obtain the reduced-form solutions (multiplying across by H^{-1} gets it into the form analyzed here). My current plan is to have you use this program to do some computations with some simple DSGE models.