OPTIMALLY RATIONAL EXPECTATIONS AND MACROECONOMICS

David Demery
Nigel W. Duck[1]

May 2002
Discussion Paper No. 02/533

Department of Economics
University of Bristol
8 Woodland Road
Bristol BS8 1TN
UK

Corresponding Author: David Demery
E-mail: David.Demery@bris.ac.uk
Telephone: +44 (0)117 928 8437
Fax: +44 (0)117 928 8577

Abstract

This paper provides an alternative to the theory of rational expectations (RE). Its central idea is that the information set on which agents will choose to condition their expectations will not, in general, include all the available information. Our alternative has many of the attractive features of RE; it emerges from an explicit choice-theoretic framework; it has wide applicability; and it can in principle explain the failure of models incorporating RE to account for the dynamics of many macroeconomic relationships.

Classification Code: E21
Key Words: Rational expectations, incomplete information, macroeconomic dynamics

[1] We are grateful to Clifford Attfield, Hedley Rees, Suro Sahay and Richard Smith for helpful comments. We are responsible for any remaining errors.

1 Introduction

Recent contributions to the New Keynesian theory of inflation have raised doubts about the validity of the full-information version of rational expectations (RE).[1] The main reason for these doubts is the model's failure to account for the observed degree of persistence in inflation. This failure is also characteristic of other areas: for example the application of RE to the permanent income model of consumption does not explain the sluggish behaviour of aggregate consumption.[2] At the same time, models of expectations formation which *can* account for the observed sluggishness, such as adaptive expectations, are widely regarded, from a theoretical perspective, as unattractive. In this paper we put forward an alternative to RE which can, in principle, account for sluggishness in aggregate variables but which is derived within a choice-theoretic framework and is not vulnerable to the objections raised against existing alternatives to RE. We call this alternative *optimally rational expectations* (ORE).

The theory we propose addresses a long-standing objection to the usual theoretical justification for RE.[3] This justification argues that since RE is based on the assumption that agents fully exploit *all available* information, no other method of forming expectations can be superior to it, and so RE is optimal. A corollary is that RE is consistent with a system in equilibrium (agents will not want to change to an inferior method of expectations formation) and accords with an equilibrium approach to modelling. Both these arguments confuse *technical* and *economic* optimality, or at least are relevant only when information is free. As Feige and Pearce (1976) and Buiter (1980) argued, the agent's decision about how to form expectations should be analysed like any other economic decision, from a choice-theoretic

[1] Mankiw (2001), for example, has voiced dissatisfaction with RE in the context of the New Keynesian model of the Phillips Curve, arguing that 'There is a simple way to reconcile the new Keynesian Phillips Curve with the data: adaptive expectations. .... Because of this, some people working in this area are now questioning the assumption of rational expectations.' In that same context, other researchers such as Galí and Gertler (1999), Galí, Gertler and Lopez-Salido (2001), and Ball (2000) and Mankiw and Reiss (2001) have begun to use alternatives to RE.

[2] This failure led many to jettison or extend the permanent income model itself but a number of contributors have, instead, relaxed the assumption of full information RE: see Pischke (1995), Goodfriend (1992), and Demery and Duck (1999, 2000). Departures from RE are becoming increasingly common in other areas of macroeconomics. Woodford (2001) allows a departure from the full-information assumption of RE in his re-working of Lucas's price-surprise model of the business cycle. And Cutler, Poterba and Summers (1991) suggest that RE also fails to explain stock-price dynamics.

[3] A separate objection to RE is that there is little direct evidence to support it, at least in its strong form. For example, after reviewing the survey evidence for RE, Conlisk (1996) concludes that these surveys 'commonly reject the unbiasedness and efficiency predictions of rational expectations' (p. 672-3); and in an earlier survey, Pesaran (1987) concluded that 'direct observations on expectations do not support the RE [Hypothesis]. It is therefore important to consider other models of expectations formation' (p. 244).

framework in which the agent weighs up the costs and benefits of forming expectations in different ways. This process will not necessarily cause the agent to select a method of expectations formation that is *technically* optimal. Furthermore, the method of expectations formation which emerges from such a framework, whether it is RE or not, is quite consistent with the usual interpretation of a system in equilibrium, i.e. a situation where agents have no incentive to change their behaviour because the marginal costs and benefits of it are in balance, rather than a situation where marginal benefits have been driven to zero.

From a choice-theoretic perspective RE's assumption that agents will condition their expectations on *all* the available information is valid when all the available information is costless.[4] The specific starting point for ORE is the view that, in most contexts, this condition will not hold. The formal acquisition cost of some information may well be trivial - information about 'the inflation rate' for example is readily available and widely publicised - but even this information can be seen as costly in the wider sense of involving time to acquire, store, update and process. And whilst information about other variables, such as the behaviour of monetary aggregates, is still relatively easy to acquire, information about more obscure variables, for example aggregate real labour income or the true profitability of a particular company, requires more substantial acquisition and processing costs. So it seems sensible to view free information as the exception rather than the rule, to view most information as having associated acquisition and other costs that are not trivial, and to assume that, when deciding how to form expectations, rational agents will improve their information set only up to the point where the marginal benefits of doing so equal the associated marginal costs.[5]

We formalise this idea in a way which is applicable to many macroeconomic, and perhaps other, contexts. We assume that the variable about which the agent forms an expectation is one of many similar variables each of which is subject to two types of shock: a macroeconomic or *common* shock; and a microeconomic or *idiosyncratic* shock specific to that variable. So, for example, the variable might be the agent's own labour income which is affected by a component that affects everyone's labour income, e.g. the state

[4]Buiter (1980) makes the point as follows: '[T]he term rational ... expectations ought to be reserved for forecasts generated by a rational, expected utility maximising decision process in which the costs of acquiring, processing and evaluating additional information are balanced against the anticipated benefits from further refinement of forecasts' (p.35).

Feige and Pearce (1976) made the same point, arguing that '[R]ational expectations models offer the theoretical appeal of greater consistency with the economist's paradigm of rational behaviour in a world of negligible information costs; however, they avoid speaking to the empirical issue of selecting a particular information set' (p.518).

[5]Galbraith (1988) made similar criticisms of RE and developed an alternative theory of expectations formation similar in spirit to ours. In his, agents apply appropriate techniques to models in which the explanatory variables are measured with error.

of the business cycle, and by factors specific to the individual. The agent can condition her forecasts of the series solely on the history of the series itself - which in many contexts will be observed costlessly as part of the agent's economic activity - or can, at a cost, choose to observe the history of other similar variables and thereby increase the accuracy of her forecasts because the extra series contain additional information on the common shock.

Within this setting, we show how the agent will form expectations conditioned on an *optimal* information set. The expectations so formed are *rational* in the sense that the agent fully exploits her chosen information set; and they are *optimal* in that the agent's choice of an information set can be seen as resulting from an explicit process of trading off the benefits of greater forecast accuracy against the costs of acquiring it. RE emerges as the special case where the chosen information set and the set of all available information coincide.

The paper is organised as follows. In the first section we show why optimising agents will generally choose to make forecasts that are less accurate than is technically possible. In the second we link the degree of forecast accuracy to the quality of the information set on which expectations are conditioned, and we derive the key result that agents will form expectations on an information set that does not include all available information. The third section examines the general characteristics of the ORE forecast errors. In the fourth we describe the main implications of ORE for macroeconomic behaviour, and in the fifth we show how, in principle, ORE models may impose testable restrictions on the data. We end with a set of conclusions.

2 Optimally Rational Expectations

Consider an agent who faces the following problem:

$$\max_{x} L = L(x, D) \tag{1}$$

where x is a choice variable (or vector of such variables) and D is an exogenous variable (or vector of exogenous variables). For convenience, in this section we omit any relevant subscripts, but note that the variables in equation (1) refer to a particular agent and time period. Assume that the agent's decision about x has to be made *prior* to the realization of D and that the agent has therefore to form an expectation of D.

Equation (1) is general enough to cover a number of familiar macroeconomic situations. For example, x could be a vector of consumption decisions over an agent's planning horizon and D the relevant resource constraint; or, in the context of a New Keynesian model, x could be viewed as the firm's price and D the firm's current and future marginal costs.

For simplicity, consider x and D as scalars. If the agent were to have full knowledge of D, the optimal value of x, x^*, would be:[6]

$$x^* = f(D) \tag{2}$$

and the resulting value of L would be:

$$L^* = L(x^*, D) \tag{3}$$

Since the agent is assumed to select x before the value of D is known, her decision will be based not on D itself but on the expectation of D. Assume that this expectation is conditioned on an information set Ω and that the agent fully exploits that information so that:

$$D = E(D \mid \Omega) + \omega \tag{4}$$

where ω is a mean-zero Gaussian white-noise error which is uncorrelated with $E(D \mid \Omega)$ and has a constant variance σ_ω^2. In what follows we shall use this variance as a measure of the accuracy of the agent's forecasting technique: the higher the value of σ_ω^2 the lower the accuracy of the forecasts.

With this expectation, the agent's optimal choice of x, x^A, will be:

$$x^A = f(E(D \mid \Omega)) \tag{5}$$

and the realized value of L will be:

$$L^A = L(x^A, D) \tag{6}$$

From a Taylor-expansion of equation (6) around x^* we write:

$$L^A = L^* + L_1(x^*, D) \cdot \left(x^A - x^*\right) + \frac{1}{2} L_2(x^*, D) \cdot \left(x^A - x^*\right)^2 + \frac{1}{6} L_3(x^*, D) \cdot \left(x^A - x^*\right)^3 + ...$$

where L_i represents the i^{th} partial derivative of the objective function with respect to x.

Ignoring derivatives of order higher than two, the difference between L^* and L^A, i.e. the loss resulting from an incorrect forecast of D, can be written as:

$$L^* - L^A = k\left(x^A - x^*\right)^2$$

where k is $\frac{1}{2} L_2(x^*, D)$ and, using equations (2), (4) and (5), we can write:

$$L^* - L^A = h(E(D \mid \Omega), \omega) \tag{7}$$

[6] A feature of the next section is that the costs of acquiring information about D enter into the agent's optimising problem. For the moment we ignore these costs.

The expected value of equation (7) is:

$$E\left(L^{*}-L^{A}\right)=\int_{-\infty}^{\infty} h\left(E\left(D \mid \Omega\right), \omega\right) \cdot g(w) d\omega \tag{8}$$

where $g(\omega)$ is the density function for ω. Writing ω as $\sigma_\omega \cdot z$, where $z \sim N(0,1)$ allows equation (8) to be written as:

$$E\left(L^{*}-L^{A}\right)=\int_{-\infty}^{\infty} h\left(E\left(D \mid \Omega\right), \sigma_\omega z\right) \cdot \frac{1}{\sigma_\omega} \phi(z) \sigma_\omega dz \tag{9}$$

where $\phi(z)$ is the density function for a standard normal variable.

As equations (4) to (9) indicate, any choice of information set can be seen as implying a particular value for σ_ω^2 and hence a particular expected deviation of utility from its maximum, L^*. Conventional optimising principles suggest that the actual information set selected will be the result of a process of weighing up the costs of acquiring a richer information set and the gain in utility from the more accurate forecasts it produces. For the moment we shall *assume* that agents can select σ_ω^2; in the next section we provide a specific model of how the choice of σ_ω^2 translates into a choice of information set. So in this section the agent's optimising problem involves calculating the marginal cost and benefits of reducing σ_ω^2.

The extra benefits of greater accuracy can be found by differentiating equation (9) with respect to σ_ω^2. This shows the loss in expected utility that results from a higher value of σ_ω^2 to be:

$$\frac{\partial E\left(L^{*}-L^{A}\right)}{\partial \sigma_\omega^2}=\int_{-\infty}^{\infty} \frac{\partial h\left(E\left(D \mid \Omega\right), \sigma_\omega z\right)}{\partial \sigma_\omega^2} \phi(z) dz$$

Since:

$$\begin{aligned}\frac{\partial h\left(E\left(D \mid \Omega\right), \sigma_\omega z\right)}{\partial \sigma_\omega^2} &= \frac{\partial h\left(E\left(D \mid \Omega\right), \sigma_\omega z\right)}{\partial(\sigma_\omega z)} \cdot \frac{\partial(\sigma_\omega z)}{\partial \sigma_\omega^2} \\ &= \frac{\partial h\left(E\left(D \mid \Omega\right), \sigma_\omega z\right)}{\partial(\sigma_\omega z)} \cdot \frac{z}{2\sigma_\omega}\end{aligned}$$

it follows from this that:

$$\frac{\partial E\left(L^{*}-L^{A}\right)}{\partial \sigma_\omega^2} > 0 \text{ if } \frac{\partial h\left(E\left(D \mid \Omega\right), \sigma_\omega z\right)}{\partial \omega} . \frac{\omega}{2\sigma_\omega^2} > 0$$

This condition is satisfied if $\operatorname{sign}(\omega) = \operatorname{sign}\left(\frac{\partial h(E(D \mid \Omega), \sigma_\omega z)}{\partial \omega}\right)$. So, the fall in utility due to a higher value of σ_ω^2 will be positive under a very weak restriction on the nature of the $h(.)$ function: that $h\left(E\left(D \mid \Omega\right), \omega\right)$ is increasing in the *absolute* value of ω. This condition is likely to be satisfied by any plausible objective function: it requires that negative and positive forecasting errors lead to losses and that the higher these errors are in *absolute terms* the greater the loss. So, the higher is σ^2_ω, the higher, in general, will be an agent's expected loss from forming imperfect expectations of D.

We shall assume that the costs of reducing σ^2_ω are positive. Subject of course to certain second-order conditions, a unique and stable optimum value of σ^2_ω will exist where a marginal increase in σ^2_ω involves an equal fall in expected utility and drop in costs (measured in units of utility).[7] In general, there is no reason why optimising agents will choose the lowest possible value of σ^2_ω. They will do so if the marginal costs of greater accuracy are zero, or if the marginal benefits always exceed marginal costs; but these are clearly special cases. In general, the value of σ^2_ω which solves the optimisation problem will be higher than its potential minimum, and agents will not choose the most accurate forecasting technique open to them. Since RE is the most accurate forecasting technique open to them, the analysis presented here suggests that they will select RE only in special circumstances. The expectations formed in the absence of these special conditions are what we call ORE, expectations which accept a degree of avoidable inaccuracy because the costs of eliminating it are too high.

3 ORE and the choice of information set

The previous section suggested that agents will form expectations which are less accurate than RE implies, but it left open the question of precisely *how* agents might choose the optimal degree of inaccuracy. In this section we provide a model of this choice that has wide applicability, especially in macroeconomics.

We assume that the agent needs to forecast the current and/or future values of the exogenous variable D_j and that there are N such series whose average value in period t, D_t, is $\frac{1}{N}\sum_{j=1}^{N} D_{j,t}$. We can express $D_{j,t}$, as:

$$D_{j,t} \equiv D_t + [D_{j,t} - D_t] \tag{10}$$

We shall assume that both D_t and $[D_{j,t} - D_t]$ are stationary and are uncorrelated with each other. Movements in $D_{j,t}$ are therefore the sum of two distinct sets of influences: *common* or *macroeconomic* influences on the one hand, and *idiosyncratic* influences on the other.[8] For example, where the agent is an individual deciding how much to consume, one of the variables to be forecast would be her labour income. This will be influenced by the state of the business cycle, clearly a macroeconomic influence, and factors related

[7] A sufficient set of conditions is that the marginal cost of acheiving greater accuracy (reducing σ^2_ω) will increase as the degree of accuracy increases, and that as σ^2_ω increases the drop in utility increases. The latter condition places conditions on the $h(.)$ function - that there is a sufficiently large increase in the loss of utility as σ^2_ω increases.

[8] It would be possible to extend equation (10) to include a third set of influences, for example regional or industry-wide influences, that are common to a subset of agents. The substantive results we draw from equation (10) would be unaffected by such extensions.

to her own particular characteristics. In what follows we shall assume that the variable to be forecast is of this type, so that j indexes both the agent who is making the forecast and the series being forecast.[9]

We represent the stationary, aggregate component, D_t, in the following invertible moving-average error form:

$$D_t = d + \alpha(L)\varepsilon_t = d + \sum_{i=0}^{T_1} \alpha_i \varepsilon_{t-i} \tag{11}$$

where d is a constant, $\alpha(L)$ is a polynomial in the lag operator with $\alpha_0 = 1$ and ε is a Gaussian white-noise error process with zero mean and variance σ_ε^2.[10]

Similarly, we represent the stationary, idiosyncratic component, $D_{j,t} - D_t$, in the following invertible moving-average error form:

$$D_{j,t} - D_t = \gamma(L)u_{j,t} = \sum_{i=0}^{T_2} \gamma_i u_{j,t-i}$$

where $\gamma(L)$ is a polynomial in the lag operator with $\gamma_0 = 1$; $u_{j,t}$ is a Gaussian white-noise error process with zero mean and variance σ_u^2. We assume that $u_{j,t}$ is uncorrelated with ε_t at all leads and lags, and is uncorrelated with $u_{j-v,t-k}$ for all values of v and k other than zero.[11] In what follows we assume that at least some of the αs are different from the equivalent γs. We also assume that $T_1 = T_2 = T$.[12]

It follows that D_{jt} can be represented as the sum of two moving-average error processes:

$$D_{j,t} = d + \alpha(L)\varepsilon_t + \gamma(L)u_{j,t} \tag{12}$$

To fix our timing convention, we assume that, for all $v \geq 0$, ε_{t+v} and $u_{j,t+v}$ *cannot* be known at t, whereas all other values of ε_{t+v} and $u_{j,t+v}$ are *feasible* elements of any agent's information set at period t.[13]

The full-information assumption of RE would imply that the histories of both the aggregate variable, $\{D_t\}$ and of $\{D_{j,t}\}$ are known to the agent.

[9] This is not strictly necessary. It would be possible to think of many agents forming expectations about the same exogenous variable, say the price of a particular good, which is influenced by both aggregate and, in this case, market-specific factors.

[10] By the Wold representation theorem covariance stationary variables have moving-average error representations. We have assumed a finite-order moving-average process but this is not essential.

[11] Equation (10) implies that $\sum_{j=1}^{N} u_{j,t-i} = 0$. Our assumption about the lack of correlation between $u_{j,t}$ and $u_{j-v,t}$ for all v requires that N is 'large'.

[12] If $T_1 \neq T_2$, then whichever is the lower-order process can always be redefined to include the required extra number of terms attached to zero coefficients.

[13] This means that no expenditure of resources would allow the agent to know the value of ε_t or $u_{j,t}$ in period t, whereas, with some finite expenditure, agents can acquire knowledge of past values of ε_t or $u_{j,t}$.

From the history of $\{D_t\}$ the agent could infer the history of $\{\varepsilon_t\}$, and, from this and the history of $\{D_{j,t}\}$, she could to infer the history of $\{u_{j,t}\}$. Conditioning her expectations on these *separate* histories of $\{\varepsilon_t\}$ and $\{u_{j,t}\}$, which constitute *all* the available information about the behaviour of $D_{j,t}$, she would form the rational expectation of the current and future values of $D_{j,t}$ as:

$$\begin{aligned} \text{If } 0 &\leq k < T: \\ E^F_{j,t} D_{j,t+k} &= d + \sum_{i=k+1}^{T} \alpha_i \varepsilon_{t+k-i} + \sum_{i=k+1}^{T} \gamma_i u_{j,t+k-i} \\ \text{Otherwise} &: \quad E^F_{j,t} D_{j,t+k} = d \end{aligned}$$

where $E^F_{j,t} D_{j,t+k}$ is the expectation of $D_{j,t+k}$ formed by agent j conditional on this full information set in period t. Note that our timing convention implies that the information set at time t does *not* include the *current* value $D_{j,t}$, but does include its history and, in this full information case, the *separate* histories of the aggregate and idiosyncratic shocks.

In contrast to the full-information case we shall assume that the aggregate variable, D_t, and hence the history of $\{\varepsilon_t\}$ *cannot* be directly observed and, as a consequence, observation of the history of $\{D_{j,t}\}$ does *not* allow observation of the history of $\{u_{j,t}\}$. Instead we shall assume that the agent can only acquire more information about the separate contributions of ε_t and $u_{j,t}$ by observing a larger number of the available N series of which D_j is one.

Whilst this is, to some extent, merely a device for modelling the acquisition of richer information sets in an analytically tractable way, there are more practical justifications for it. First, most published macroeconomic series, the most obvious source of a direct observation on D_t, are best seen as *indicators* of the true corresponding economic variable. For example, most countries publish several different money supply measures, all of which are almost certainly imperfect empirical counterparts to the theoretical construct, the quantity of money. From this perspective the published variable, D^P_t, can be seen as equal to the true variable, D_t, plus measurement error. So we could write $D^P_t = D_t + \phi(L)\zeta_t$ where ζ_t is a Gaussian white-noise error term and $\phi(L)$ is a polynomial in the lag operator with $\phi_0 = 1$. In this case, to forecast $D_{j,t}$, the agent needs to forecast $D_{j,t} - D^P_t$ and the agent's problem has exactly the same form as equation (12). Rather than introduce this complication explicitly, we assume that the aggregate variable, D_t, cannot be directly observed. Even, if the published series provided in principle a perfect empirical counterpart to D_t, there would be the inevitable data revisions and publication delays which would, in practice, make D_t (especially its most recent values) difficult or even impossible to observe.

As an example of how a rational agent will form expectations when the *separate* histories of $\{\varepsilon_t\}$ and $\{u_{j,t}\}$ are *not* part of her information set, consider first an agent whose information set consists solely of the history of $\{D_{j,t}\}$ itself. In many settings this is a natural assumption or at least a natural starting point. If $D_{j,t}$ refers to a variable specific to the agent, her own labour income for example, then the history of $\{D_{j,t}\}$ will *necessarily* be observed by her from her participation in the appropriate economic activity, and is therefore certain to be part of her information set.[14] Knowledge of the separate components of $D_{j,t}$ is not, in the same sense, freely available; it requires knowledge of *aggregate* labour income which, in the absence of a direct observation, can only be observed from observation of all N similar series. Of course, the agent may choose to acquire such information but for the moment we assume that she does not.

An information set consisting solely of the history of $\{D_{j,t}\}$ could equally well be seen as consisting of the history of a white-noise error term $\left\{\eta_{j,t}\right\}$, where $\eta_{j,t}$ is defined by the re-parameterisation of equation (12) as:

$$D_{j,t} = d + \theta(L)\eta_{j,t} = d + \sum_{i=0}^{T} \theta_i \eta_{j,t-i}$$

and where θ_i is a function of the α_is, the γ_is, and the two variances, σ_ε^2, and σ_u^2.[15]

If the agent were fully to exploit *this* information set, i.e. were to condition her forecast on the history of $\left\{\eta_{j,t}\right\}$, the agent's expectations would be:

$$\begin{aligned} \text{If } 0 &\leq k < T\text{:} \\ E^1_{j,t}D_{j,t+k} &= d + \sum_{i=k+1}^{T} \theta_i \eta_{j,t+k-i} \\ \text{Otherwise} &: \quad E^1_{j,t}D_{j,t+k} = d \end{aligned}$$

where the notation $E^1_{j,t}$ denotes agents j's expectations based on the information set that assumes knowledge of the history of only one series, D_j.

Assume now a more general case where the typical agent's information set, whilst still incomplete in the sense of not containing the *separate* histories of $\{\varepsilon_t\}$ and $\{u_{j,t}\}$, contains the separate histories of $\{D_{j,t}\}$ and, at some cost, *one* other series, for example $\{D_{j+1,t}\}$. The two observed histories, *each taken on its own*, amount to observations on $\{d + \alpha(L)\varepsilon_t + \gamma(L)u_{j,t}\}$ from

[14] Even though the acquisition costs of the history of $\{D_{j,t}\}$ might be costless, the storage and processing costs might not be. So even in this case it may be inaccurate to regard the information as costless. However, the history of $\{D_{j,t}\}$ will, for agent j, be sunstantially *less* costly to acquire than the costs of acquiring the history of any other series.

[15] See Hamilton (1994, pp. 102-107). In the simple MA(1) case, θ_1 is the invertible solution to the quadratic equation $A\theta_1^2 + B\theta_1 + C$ where $A = C = \alpha_1 + \gamma_1\sigma_u^2/\sigma_\varepsilon^2$; and $B = -\left[\left(1+\alpha_1^2\right) + \left(1+\gamma_1^2\right)\sigma_u^2/\sigma_\varepsilon^2\right]$. As $\sigma_u^2/\sigma_\varepsilon^2 \Rightarrow \infty$ $\theta_1 \Rightarrow \gamma_1$; and as $\sigma_u^2/\sigma_\varepsilon^2 \Rightarrow 0$ $\theta_1 \Rightarrow \alpha_1$.

$\{D_{j,t}\}$, and $\{d+\alpha(L)\varepsilon_t+\gamma(L)u_{j+1,t}\}$ from $\{D_{j+1,t}\}$.[16] However, *taken together*, they also amount to an observation on the history of the mean of two series, $\{\overline{D}_{2,j,t}\}$, where $\overline{D}_{2,j,t} \equiv \frac{D_{j,t}+D_{j+1,t}}{2}$, and:

$$\overline{D}_{2,j,t} = d+\alpha(L)\varepsilon_t+\gamma(L)\overline{u}_{2,j,t} = d+\sum_{i=0}^{T}\alpha_i\varepsilon_{t-i}+\sum_{i=0}^{T}\gamma_i\overline{u}_{2,j,t-i}$$

$\overline{u}_{2,j,t}\left(\equiv \frac{u_{j,t}+u_{j+1,t}}{2}\right)$ is a white noise error with variance $\sigma^2_{u_2}\left(=\frac{1}{2}\sigma^2_u\right)$. From the re-parameterising of this expression it is clear that, from the *joint observation* of the histories of $\{D_{j,t}\}$ and $\{D_{j+1,t}\}$, agents obtain the history of the white noise error term, $\overline{\eta}_{2,j,t}$, defined by:

$$\overline{D}_{2,j,t} = d+\theta_2(L)\overline{\eta}_{2,j,t} = d+\sum_{i=0}^{T}\theta_{2,i}\overline{\eta}_{2,j,t-i}$$

where $\theta_2(L)$ is a polynomial in the lag operator with $\theta_{2,0}=1$, and where $\sigma^2_{\eta_2}$, the variance of $\overline{\eta}_{2,t}$, and the elements of $\theta_2(L)$s are functions of the αs, the γs, and the two variances σ^2_ε, and $\sigma^2_{u_2}$.

It is straightforward to generalize this result to the case where an agent's information set contains the history of $\{D_{j,t}\}$ and the histories of $(Y-1)$ other such series $\{D_{j+1,t}...D_{j+Y-1,t}\}$, each of which is costly to acquire. In our framework, the information provided by any one other series is, for agent j, as good as the information provided by any other. We assume then that each series (other than $D_{j,t}$ itself) has an equal probability of being part of agent j's information set.[17] The use of the j subscript when referring to averages emphasises that no two agents share the same information set other than by chance, though we do assume in what follows that the optimal value of Y is common to all. In this case an agent observes the history of the white-noise error, $\overline{\eta}_{Y,j.t}$, where:

$$\begin{aligned}\overline{D}_{Y,j,t} &= d+\alpha(L)\varepsilon_t+\gamma(L)\overline{u}_{Y,j,t}\\ &= d+\theta_Y(L)\overline{\eta}_{Y,j,t} = d+\sum_{i=0}^{T}\theta_{Y,i}\overline{\eta}_{Y,j,t-i}\end{aligned} \tag{13}$$

[16]Note that we assume that σ^2_u and the parameters d, α_i and γ_i are common to all agents.

[17]In reality, it is more likely that agents will 'network' and that the probability any other series has of being included in agent j's information set will depend upon (say) its geographical or occupational proximity to agent j. In this 'networking' case, groups of agents will share the same information set, each being informed of the others' Ds and their histories. Such informal channels are likely to be important sources of low-cost information and often integral to the agent's participation in economic activity. As long as these clusters of information pooling are small relative to the economy at large, our conclusions are unaffected. In the interests of simplicity we assume that each agent j obtains information from a random draw of $Y-1$ realisations of D. This means that no two agents will share the same information set other than by chance.

$\overline{D}_{Y,j,t} = \frac{1}{Y}\sum_{i=0}^{Y-1} D_{j+i,t}$; $\overline{u}_{Y,j,t} \equiv \frac{1}{Y}\sum_{i=0}^{Y-1} u_{j+i,t}$; $\theta_Y(L)$ is a polynomial in the lag operator with $\theta_{Y,0} = 1$; and the variance of $\overline{\eta}_{Y,j,t}$, $\sigma^2_{\eta_Y}$, and the elements of $\theta_Y(L)$s are functions of the αs, the γs, and the two variances σ^2_ε, and $\sigma^2_{u_Y}$ $\left(= \frac{1}{Y}\sigma^2_u\right)$.

From equations (13) and (12) it follows that, for this agent, $D_{j,t}$ can be represented as the sum of *two* white-noise error processes whose histories are *separately* observed, $\left\{\overline{\eta}_{Y,j,t}\right\}$ and $\{u_{j,t} - \overline{u}_{Y,j,t}\}$:

$$\begin{aligned} D_{j,t} &= d + \theta_Y(L)\overline{\eta}_{Y,j,t} + \gamma(L)(u_{j,t} - \overline{u}_{Y,j,t}) \\ &= d + \sum_{i=0}^{T}\theta_{Y,i}\overline{\eta}_{Y,j,t-i} + \sum_{i=0}^{T}\gamma_i(u_{j,t-i} - \overline{u}_{Y,j,t-i}) \end{aligned} \quad (14)$$

To repeat, the history of $\left\{\overline{\eta}_{Y,j,t}\right\}$ is observed from the *joint* observation of Y series; and the history of $\{u_{j,t} - \overline{u}_{Y,j,t}\}$ is the *extra* information obtained from observing the *separate* history of the jth series.

It follows from equation (14) that an agent who observes $\{D_{j,t}...D_{j+Y-1,t}\}$, the *separate* Y histories, and fully exploits that information, will form expectations of the current and future values of $D_{j,t}$ as follows:

$$\begin{aligned} &\text{If } 0 \leq k < T: \\ E^Y_{j,t}D_{j,t+k} &= d + \sum_{i=k+1}^{T}\theta_{Y,i}\overline{\eta}_{Y,j,t+k-i} + \sum_{i=k+1}^{T}\gamma_i(u_{j,t+k-i} - \overline{u}_{Y,j,t+k-i}) \\ \text{Otherwise} &: E^Y_{j,t}D_{j,t+k} = d \end{aligned} \quad (15)$$

where the notation $E^Y_{j,t}$ defines agent j's expectations conditioned on an information set consisting of Y histories. Once again, 'fully exploits' means that no element in the histories of, in this case, $\left\{\overline{\eta}_{Y,j,t}\right\}$ and $\{u_{j,t} - \overline{u}_{Y,j,t}\}$ can be used to reduce the agent's forecast error.[18]

Equation (15) begins to make operational the ideas developed in the previous section. Agents have a choice about how to form their expectations of $D_{j,t+k}$. Specifically, they can choose the information set on which to condition those expectations: they can choose to condition them on a single series $\{D_{j,t}\}$ or on an information set that includes up to N such series. Hence the trade off established between accuracy and cost in the previous section translates into a trade-off between the extra cost and greater accuracy that result from an agent having a larger number of series her information set. Note that from this perspective RE can seen as a special case of ORE where the number of series included in the information set equals the number of observable series: as $Y \Rightarrow N$, $(u_{j,t} - \overline{u}_{Y,j,t}) \Rightarrow u_{j,t}$ for all t, the history of $\{u_{j,t}\}$ is revealed, and $\theta_Y(L)\overline{\eta}_{Y,j,t} \Rightarrow \alpha(L)\varepsilon_t$. Hence the agent observes the *separate* histories of $\{\varepsilon_t\}$ and $\{u_{j,t}\}$ and is able to condition her expectations on the 'available' information set.

[18]We demonstrate in Appendix A that the forecast error from equation (15) is white noise.

4 Characteristics of ORE forecast errors

To establish the main characteristics of ORE and compare them with the characteristics of RE, we define the one-period ahead forecast error of $D_{j,t}$ as $D_{j,t} - E_{j,t}D_{j,t}$. Under RE, expectations are conditioned on the full information set, i.e. the *separate* histories of $\{\varepsilon_t\}$ and $\{u_{j,t}\}$. The one-period-ahead forecast error, $R^F_{j,t}$, will equal:

$$R^F_{j,t} = \varepsilon_t + u_{j,t}$$

Since this is the sum of two uncorrelated white-noise errors, it too is white-noise. So, in this case, the typical agent's one-period-ahead forecast errors will, on average, be zero and they will show no pattern. Since the separate histories of $\{\varepsilon_t\}$ and $\{u_{j,t}\}$ represent the total available information, the forecast errors will be unpredictable from *any* available information. The variance of the forecast errors, $\left(\sigma^2_\varepsilon + \sigma^2_u\right)$, is simply the inherently unpredictable component of the variable and hence no other forecasting technique could produce a lower variance.

If, on the other hand, expectations were conditioned on an information set consisting of the histories of a sub-set of Y histories, $\{D_{1t}...D_{Yt}\}$, and not the *separate* histories of $\{\varepsilon_t\}$ and $\{u_{j,t}\}$, then the one-period-ahead forecast error, $R_{Y,j,t}$, would, from equation (15), equal:

$$R_{Y,j,t} = \overline{\eta}_{Y,j,t} + u_{j,t} - \overline{u}_{Y,j,t} \tag{16}$$

As we show in Appendix A, this error too is white noise. So, conditioning on this incomplete information set, an agent's one-period-ahead forecast errors will also on average be zero and they too will appear *to the agent* to be patternless and unpredictable *from any information contained in the agent's information set.*[19] The agent, in this case too, is fully exploiting her chosen information set: the forecast-error variance will in this case be $\sigma^2_{\eta_Y} + \left[\frac{Y-1}{Y}\right]\sigma^2_u$ and no other forecasting technique that uses the same information set will produce a forecast error-variance lower than this.

The actual forecast error variance will depend upon the number of series the agent includes in her information set. The addition of another series to an existing information set implies in increase in information and hence an improvement in the accuracy of the resulting forecasts until the variance of the forecast error is driven to its minimum value, $\left(\sigma^2_\varepsilon + \sigma^2_u\right)$, the value when all N series are observed and the agent can perfectly distinguish the histories of the common and idiosyncratic components.[20]

[19] However, *to an outside observer,* with information on the *separate* histories of $\{\varepsilon_t\}$ and $\{u_{j,t}\}$, the agent's forecast error would not be patternless and entirely unpredictable; the value of $\overline{\eta}_{Y,j,t}$ is to some extent predictable from knowledge of lagged values of ε_t and $u_{j,t}$, knowledge which the agent is assumed not to have. This may be why many direct tests of RE fail.

[20] Appendix B elaborates on this point.

The precise way in which the forecast-error variance drops as more series are included in the information set depends upon the parameters of the $\alpha(L)$ and $\gamma(L)$ processes and on $\sigma_u^2/\sigma_\varepsilon^2$. In Figures 1 and 2 and Table 1 we illustrate this relationship for two sets of MA(1) processes with σ_ε^2 normalised to one. In the first set, α_1 and γ_1 are very different; in the second they are quite similar. The figures show how $\left(\sigma_{\eta_Y}^2 + \left[\frac{Y-1}{Y}\right]\sigma_u^2\right)$ varies as Y, the number of series observed, increases from 1 to 100. The table shows the precise values of $\left(\sigma_{\eta_Y}^2 + \left[\frac{Y-1}{Y}\right]\sigma_u^2\right)$ for selected values of Y.

The figures and table illustrate a number of points. First, whilst an additional series always leads to an increase in accuracy, the marginal increase in accuracy falls sharply as the number of series increases. Secondly, given Y, σ_u^2 and σ_ε^2, it is more difficult to forecast accurately when α_1 and γ_1 are very different. The reason is that, if the two *processes* are similar, there is less to be gained from distinguishing the two *shocks*; in the limit, if the two processes are identical there is *nothing* to be gained from distinguishing between them, the two shocks can be treated as a single shock. Thirdly, increases in σ_u^2 (or σ_ε^2) will increase the inaccuracy of forecasts, though this will be offset to some extent by changes in θ_1. For example, a larger value for, say, σ_u^2, makes $D_{j,t}$ inherently more difficult to predict and so raises the value of the forecast error variance; but it also implies a higher value for $\sigma_u^2/\sigma_\varepsilon^2$. This in turn will imply a value for θ_1 closer to γ_1.[21] The reparameterised equation will therefore predict the movements in $D_{j,t}$ due to movements in $u_{j,t}$ more closely than it would if θ_1 had remained constant. By the same argument, a larger value of σ_ε^2, whilst making $D_{j,t}$ inherently more difficult to predict, will lower $\sigma_u^2/\sigma_\varepsilon^2$ and push the value of θ_1 towards α_1. The reparameterised equation will therefore predict the movements in $D_{j,t}$ due to movements in ε_t more closely than it would if θ_1 remained constant. This effect suggests that the increase in inaccuracy caused by increases in either σ_u^2 or σ_ε^2 may well be quite modest.

More generally, the figures and table suggest that the optimal value of Y will depend upon the particular characteristics of the process determining $D_{j,t}$.[22] They suggest that it is quite possible that the marginal benefits of more information fall quickly to very low levels. If, as the information set increases in size, the marginal costs of processing, storing and updating another series rise, then even a small number of series might deliver the optimal degree of forecast accuracy and the agent might never feel it necessary to observe the behaviour of the aggregate variable.[23] It is even possible for the

[21] As footnote (14) explains, in this case as $\sigma_u^2/\sigma_\varepsilon^2 \Rightarrow \infty$ $\theta_1 \Rightarrow \gamma_1$; and as $\sigma_u^2/\sigma_\varepsilon^2 \Rightarrow 0$ $\theta_1 \Rightarrow \alpha_1$.

[22] The same general pattern was observed in cases where we assumed MA processes of a higher order.

[23] In their application of ORE to a New Keynesian model of inflation, Demery and Duck (2001) find that the loss of profit to the firm from not observing the two shocks separately is likely to be very low even for the case where $Y = 1$. In a study of the Permanent Income

optimal value of Y to be less than the number actually observed through the agent's participation in economic activity. The opposite - that the optimal degree of accuracy requires the inclusion of the complete set of series so that agents form RE - cannot, of course, be ruled out but within the framework developed here it appears as a special case.

5 Macroeconomic implications of ORE

Once one allows for the differences in information sets, RE and ORE have apparently similar characteristics but there are in fact important differences between them, especially in their implications for aggregate variables. These differences arise from the fact that whilst R_t^F, the RE economy-wide or average forecast error, is white noise, the ORE equivalent, $R_{Y,t}$, is not.

To see this note that, given full information, R_t^F is the sum of $R_{j,t}^F$ over j, divided by N. Since by definition, the terms in $u_{j,t}$ sum to zero across j, this will be ε_t, and since ε_t is defined to be a white-noise error, the *average* forecast error is also a white-noise error. The ORE average forecast error is the sum of equation (16) over j divided by N. Our assumption about the agents' choice of which other series to include in their information sets (i.e. that each series, other than their own, has an equal probability of being included), and the assumption that the terms in $u_{j,t}$ sum to zero across j, together imply that the second and third terms in equation (16) sum to zero. From equation (11) and an aggregation of (14) the following expression for the average value of $\overline{\eta}_{Y,t}$ can derived:

$$\overline{\eta}_{Y,t} = \frac{\alpha(L)}{\theta_Y(L)}\varepsilon_t \tag{17}$$

It is clear from this that unless $Y = N$ (in which case $\alpha(L) = \theta_Y(L)$, and $\overline{\eta}_{Y,t} = \varepsilon_t$)[24] the *aggregate* equivalent of $\overline{\eta}_{Y,j,t}$, $\overline{\eta}_{Y,t}$, is *not* a white noise error - it will be predictable from its own history. Of course this is of no use to any individual agent since, unless $Y = N$, they do *not* observe $\overline{\eta}_{Y,t}$, it is not part of their information set, but this does have significant macroeconomic implications.

To illustrate these implications, assume the following simple linear form for equation (5):

$$x_{j,t}^A = \lambda_0 + \lambda_1 E_{j,t}^Y D_{j,t} \tag{18}$$

Hypothesis Pischke (1995) showed that even in the case where $Y = 1$ the loss of utility suffered by an agent from not observing aggregate labour income was very small. In such cases, even a small cost might prevent the agent from observing any published series on the history of $\{D_t\}$.

[24] This condition will also be true for $Y \neq N$ if the $\alpha(L) = \gamma(L)$ so that the distinction between the common and idiosyncratic shocks is unimportant.

Assume also that agent j chooses to condition her expectations on an information set containing the histories of $\{D_{j,t}\}$ and $Y-1$ other such series. So she forms expectations using equation (15) which we repeat here for $k=0$:

$$E^Y_{j,t}D_{j,t} = d + \sum_{i=1}^{T}\theta_{Y,i}\overline{\eta}_{Y,j,t-i} + \sum_{i=1}^{T}\gamma_i(u_{j,t-i} - \overline{u}_{Y,j,t-i}) \tag{19}$$

The aggregate or average value of $x^A_{j,t}$, x^A_t, can be written:

$$x^A_t = \lambda_0 + \lambda_1\left(d + \sum_{i=1}^{T}\theta_{Y,i}\overline{\eta}_{Y,t-i}\right) \tag{20}$$

From equation (17) it follows that we can write:

$$x^A_t = \lambda_0 + \lambda_1\left(d + \sum_{i=1}^{T}\theta_{Y,i}\frac{\alpha(L)\varepsilon_{t-i}}{\theta_Y(L)}\right) \tag{21}$$

or:

$$\begin{aligned} x^A_t &= \theta_Y(L)\lambda_0 + \lambda_1\left(\theta_Y(L)d + \sum_{i=1}^{T}\theta_{Y,i}\alpha(L)\varepsilon_{t-i}\right) \\ &\quad -\theta_{Y,1}x^A_{t-1} - \theta_{Y,2}x^A_{t-2} - \ldots - \theta_{Y,T}x^A_{t-T} \end{aligned} \tag{22}$$

Two features of equation (22) are important. First, ORE implies quite different dynamics for x^A_t than does RE. RE implies that $\alpha(L)\varepsilon_t = \theta_N(L)\overline{\eta}_{N,t}$ so that equation (22) simplifies to:

$$x^A_t = \lambda_0 + \lambda_1\left(d + \sum_{i=1}^{T}\alpha_i\varepsilon_{t-i}\right) \tag{23}$$

Equation (23) is an $MA(T)$ process whereas equation (22) is an $ARMA(T, 2T)$. So a key effect of agents using an incomplete information set is that the dynamic behaviour of x^A_t is changed: specifically the response of x^A_t to a common shock is more drawn out.

The second notable, but less obvious, feature of equation (22) is that the response of the aggregate, x^A_t, to ε_t may give a quite misleading impression of the response of $x^A_{j,t}$, to a change in $E_{j,t}D_{j,t}$. Assume, for illustrative purposes, that each agent's information set consists solely of the history of her own $D_{j,t}$, so $Y=1$. Assume also that $T=1$, $\alpha_1 = 0.9$, $\gamma_1 = -0.8$ and $\sigma^2_u/\sigma^2_\varepsilon$ is sufficiently high for $\theta_{Y,1}$ to be closer to γ_1 than to α_1, say -0.2. Then, from equation (22) the aggregate variable, x^A_t, will follow the process:

$$x^A_t = 0.8\lambda_0 + 0.8\lambda_1 d - 0.2\lambda_1\varepsilon_{t-1} - 0.18\lambda_1\varepsilon_{t-2} + 0.2x^A_{t-1}$$

If, on the other hand, each agent were to observe the composite and idiosyncratic shocks separately, x^A_t would follow the process:

$$x^A_t = \lambda_0 + \lambda_1 d + \lambda_1\alpha_1\varepsilon_{t-1}$$

With an incomplete information set, not only would the behaviour of x_t^A be more drawn out, but the response of x_t^A to $\alpha_1\varepsilon_{t-1}$, i.e. to what would conventionally be measured as the change in the typical agent's expectation of $D_{j,t}$, would appear to have the wrong sign. Thus, aggregate data would give a highly misleading picture of the behaviour of the agent.

The reason for both these features is the same. If the idiosyncratic and common shocks were *separately* observed, agents would react to the shocks separately, correctly and immediately. On aggregation, the effects of the idiosyncratic shocks would be averaged out to zero and the aggregate variable would show the true effect of the common shock on individual behaviour. But if the shocks were not separately observed by the agent, the effects of the two shocks on individual behaviour would be conflated, and, if the MA processes for the two shocks were different, this conflation would induce individuals to react quite incorrectly to a common shock in the belief that it was idiosyncratic. On aggregation, this incorrect response will appear as the aggregate response to the common shock so it would appear from the behaviour of the aggregate variable as if individuals were responding to the common shock perversely.

For example, given the MA processes assumed above, a positive idiosyncratic shock in period $t-1$ would imply a lower value for $E_{j,t}D_{j,t}$ and hence induce a fall in $x_{j,t}^A$. So the typical agent, mistaking a positive value for ε_{t-1} for a positive idiosyncratic shock, would reduce $x_{j,t}^A$. Given the high value of $\sigma_u^2/\sigma_\varepsilon^2$, the typical agent *would* generally mistake a positive value for the common shock as a positive value for an idiosyncratic shock, and so would react to positive movements in the common shock by reducing $x_{j,t}^A$. The relationship between x_t^A and ε_{t-1} would therefore appear perverse. In the end, when the effects of the shock on the observed series $\{D_{j,t}\}$ have been lost, and in the absence of any other shocks, the agent will have responded 'correctly' and be in the same position as an agent who observed the initial shock separately; but when agents do not observe the shocks separately, this 'correct' response is achieved more gradually. Although we have illustrated these properties on the assumption that the agent's information set consists solely of the history of her own $D_{j,t}$, they also apply to cases where $1 < Y < N$.

There will always be a tendency for these two features to occur when agents choose to condition their expectations on an incomplete information set, but they will be more severe the greater the differences between the processes driving the common and idiosyncratic components of $D_{j,t}$, and the greater the variance of the idiosyncratic component. The first problem suggests that at least some of the inertia we observe in macroeconomic data can be viewed not as the result of (say) irrational behaviour, or the presence of adjustment costs or the stickiness of prices, but as the result of the perfectly rational behaviour of agents faced with costs of acquiring information. Imperfections in their chosen information set introduce a degree of cloudiness into agents' perceptions of the shocks affecting them, and this slows down their response. The framework we have developed is similar in spirit to that of Akerlof and Yellen (1985) and others who argued that departures from rationality ('near-rationality') involve second-order losses but have first-order macroeconomic consequences. In our case there are no departures from rationality - the small second-order losses with first-order macroeconomic consequences arise from the rational use of incomplete information sets.

6 Testing models incorporating ORE

An attractive feature of RE models is that they impose testable restrictions on the data. So too will models which incorporate ORE, though, since RE is always a special case of ORE, these restrictions are inevitably less severe. We illustrate this point using the same simple framework used above.[25] Consider the model consisting of equations (19) to (23) and the aggregate version of equation (12) with all constant terms set at zero for simplicity. After some straightforward manipulation we can write the model in the form:

$$D_t = \alpha(L)\varepsilon_t \tag{24}$$

$$x_t^A = \lambda_1 \theta^*(L) D_{t-1} - \theta^*(L) x_{t-1}^A \tag{25}$$

$$x_t^A = \lambda_1 \theta^*(L) \alpha(L) \varepsilon_{t-1} - \theta^*(L) x_{t-1}^A \tag{26}$$

where $\theta_i^* = \theta_{i+1}$.

One way to test the model would be to estimate equation (25) and test the within-equation restrictions, that the coefficients on each lag of D_t should equal the coefficients on the equivalent lag of x_t^A times $-\lambda_1$. There will be $T-1$ such restrictions where T is the order of $\theta(L)$. However, whilst this is a joint test of equation (18) and ORE against some other model, it does not allow us to discriminate between the RE version of the model and the ORE version. The RE restrictions are simply: $\theta^*(L) = \alpha^*(L)$, where $\alpha_i^* = \alpha_{i+1}$. They imply an additional T cross-equation restrictions on the data. Under these restrictions equation (26) can be written more simply as:

$$x_t^A = \lambda_1 \alpha^*(L) \varepsilon_{t-1} \tag{27}$$

The RE special case can be tested by estimating restricted and unrestricted versions of equations (24) and (27). So one strategy for investigating the nature of expectations in this simple model would involve these steps:

[25] It would be straightforward to extend his analysis to more complex cases such as the case where $x_{j,t} = \sum_{i=0}^{\infty} \lambda_i E_{j,t} D_{j,t+i}$.

(i) estimate equations (24) and (27) and test the model's cross-equation restrictions;
(ii) if (i) leads to a rejection, estimate equation (25) and test its within equation restrictions.

These procedures only apply to the simple model being considered. If equation (18) were to involve lags of $x_{j,t}^{A}$ the restrictions would be very different. Our aim here was to show that despite being a more general model of expectations-formation, ORE can impose testable restrictions on observable data.[26]

7 Conclusions

Many macroeconomic models consist of three components: (i) a representative agent who is assumed to choose the level of an activity to maximise an objective function subject to a set of constraints; (ii) a mechanism by which agents form any expectations required as part of that maximization process; (iii) the aggregation of each individual's selected level of activity to the equivalent macroeconomic variable. For some considerable time it has become almost automatic to use RE as the second component. In this paper we have suggested that, in general, RE will be inconsistent with the principles governing the first component of this process, and we have put forward an alternative which has important implications for the third.

Our alternative model includes RE as a special case but predicts that, in general, agents will accept a greater degree of inaccuracy in their forecasts than RE implies. In a specific version of this model we relate the degree of forecast accuracy to the comprehensiveness of the information set. The resulting expectations have some characteristics similar to those of RE: they produce errors that are appear patternless to the agent making them; given the selected information set, the variance of their forecast errors are as low as the variance of any other method applied to the same information set; they are quite consistent with general equilibrium in that no agent has an incentive to improve upon them; and they are capable of generating testable restrictions on the behaviour of observable data. But they also differ from RE in their implications for macroeconomic variables: they suggest that the method of expectations formation itself can induce inertia in macroeconomic behaviour; and they imply that under certain circumstances the behaviour of macroeconomic variables will be a misleading guide to the behaviour of individual behaviour.

In principle then, ORE is capable of accounting for those empirical failures of models incorporating RE which were the main spur to our development of an alternative theory of expectations formation. A small number

[26] Demery and Duck (2001) show how the application of ORE to a New Keynesian model of inflation leads to testable restrictions which they fail to reject using US and UK data.

of studies using a form of ORE have suggested that they might account for such failures in practice. Whether models incorporating ORE are more widely successful is a matter for future research.

8 Appendices

8.1 Appendix A: Proof that $R_{Y,j,t}$ is white noise

From equation (13), we have:

$$\overline{\eta}_{Y,j,t} = \frac{\alpha(L)\varepsilon_t + \gamma(L)\overline{u}_{Y,j,t}}{\theta_Y(L)}$$

which can be rewritten as:

$$\begin{aligned}\overline{\eta}_{Y,j,t} \;\; &= \;\; \alpha(L)\varepsilon_t + \gamma(L)\overline{u}_{Y,j,t} - \theta_{Y,1}\overline{\eta}_{Y,j,t-1} - \theta_{Y,2}\overline{\eta}_{Y,j,t-2} \\ &\qquad - .... - \theta_{Y,q}\overline{\eta}_{Y,j,t-q}\end{aligned}$$

By definition:

$$\overline{u}_{Y,j,t} \equiv \frac{1}{Y}\sum_{i=0}^{Y-1} u_{j+i,t}$$

The $\theta_{Y,i}$s are defined so that the covariance terms $\sigma_{\overline{\eta}_{Y,t}\overline{\eta}_{Y,t-s}} = 0$ for all $s \neq 0$.

Define the two variables:

$$y_t = \overline{\eta}_{Y,j,t} + u_{jt} - \overline{u}_{Y,j,t}$$

$$y_{t-s} = \overline{\eta}_{Y,j,t-s} + u_{j,t-s} - \overline{u}_{Y,j,t-s}$$

Then, since both $u_{j,t}$ and $\overline{u}_{Y,j,t}$ are white noise, it follows that the covariance between y_t and y_{t-s} is:

$$\sigma_{yy_{-s}} = \sigma_{u_{j,t-s}\overline{\eta}_{Y,t}} - \sigma_{\overline{u}_{Y,t-s}\overline{\eta}_{Y,t}}$$

Given our assumptions, the two covariances on the right-hand side of this expression can be written as:

$$\begin{aligned}\sigma_{u_{jt-s}\overline{\eta}_{Y,t}} \;\; &= \;\; \gamma_s\sigma_{uu_Y} - \theta_{Y,1}\sigma_{u_{jt-s}\overline{\eta}_{Y,t-1}} - \theta_{Y,2}\sigma_{u_{jt-s}\overline{\eta}_{Y,t-2}} - .... \\ &\qquad -\theta_{Y,s}\sigma_{u_{jt-s}\overline{\eta}_{Y,t-s}}\end{aligned}$$

and:

$$\begin{aligned}\sigma_{u_{Y,t-s}\overline{\eta}_{Y,t}} \;\; &= \;\; \gamma_s\sigma^2_{u_Y} - \theta_{Y,1}\sigma_{u_{Y,t-s}\overline{\eta}_{Y,t-1}} - \theta_{Y,2}\sigma_{u_{Y,t-s}\overline{\eta}_{Y,t-2}} - .... \\ &\qquad -\theta_{Y,s}\sigma_{u_{Y,t-s}\overline{\eta}_{Y,t-s}}\end{aligned}$$

since the covariances $\sigma_{u_{jt-s}\overline{\eta}_{Y,t-(s+k)}} = \sigma_{u_{Y,t-s}\overline{\eta}_{Y,t-(s+k)}} = 0$ for $k > 0$.

The equivalent terms on the right-hand side of these two expressions are the same and hence $\sigma_{u_{j,t-s}\overline{\eta}_{Y,jt}} = \sigma_{u_{Y,t-s}\overline{\eta}_{Y,t}}$ for all s and so $\sigma_{yy_{-s}} = 0$. To prove this, first note that since $\sigma_{uu_Y} \equiv \frac{\sigma^2_u}{Y}$ and $\sigma^2_{u_Y} \equiv \frac{Y\sigma^2_u}{Y^2}$:

$$\sigma_{uu_Y} = \sigma^2_{u_Y}$$

From this it follows that:

$$\sigma_{u_{jt-s}\overline{\eta}_{Y,t-s}} = \sigma_{u_{Y,t-s}\overline{\eta}_{Y,t-s}}$$

$$\sigma_{u_{j,t-s}\overline{\eta}_{Y,t-(s-1)}} = \gamma_1 \sigma_{u_Y}^2 - \theta_{Y,1}\sigma_{u_{j,t-s}\overline{\eta}_{Y,t-s}}$$

$$\sigma_{u_{Y,t-s}\overline{\eta}_{Y,t-(s-1)}} = \gamma_1 \sigma_{u_Y}^2 - \theta_{Y,1}\sigma_{u_{Y,t-s}\overline{\eta}_{Y,t-s}}$$

Both terms on the right-hand side of the second of these three expressions equal their equivalents on the right-hand side of the third. So:

$$\sigma_{u_{j,t-s}\overline{\eta}_{Y,t-(s-1)}} = \sigma_{u_{Y,t-s}\overline{\eta}_{Y,t-(s-1)}}$$

A similar argument applies to all the other relevant covariances in the expressions for $\sigma_{u_{j,t-s}\overline{\eta}_{Y,t}}$ and $\sigma_{u_{Y,t-s}\overline{\eta}_{Y,t}}$. Hence both expressions will equal zero.

8.2 Appendix B: Proof that a richer information set lowers the forecast-error variance.

Each information set has associated with it (a) a forecast error which, in the main text, we have denoted by $R_{Y,j,t} = \overline{\eta}_{Y,j,t} + u_{j,t} - \overline{u}_{Y,j,t}$; and (b) an informational history which can be expressed as the separate histories of the two series, $\{\overline{\eta}_{Y,j,t}\}$ and $\{u_{j,t} - \overline{u}_{Y,j,t}\}$.

Compare now the cases where $Y = q$ and $Y = q+1$. In the former case the forecast error is $\overline{\eta}_{q,j,t} + u_{j,t} - \overline{u}_{q,j,t}$, and the available histories are those of $\{\overline{\eta}_{q,j,t}\}$ and $\{u_{j,t} - \overline{u}_{q,j,t}\}$ where $\theta_q(L)\overline{\eta}_{q,j,t} = \alpha(L)\varepsilon_t + \gamma(L)\overline{u}_{q,j,t}$. In the latter case the forecast error is $\overline{\eta}_{q+1,j,t} + u_{j,t} - \overline{u}_{q+1,j,t}$ and the available histories are those of $\{\overline{\eta}_{q+1,j,t}\}$ and $\{u_{j,t} - \overline{u}_{q+1,j,t}\}$ where $\theta_{q+1}(L)\overline{\eta}_{q+1,j,t} = \alpha(L)\varepsilon_t + \gamma(L)\overline{u}_{q+1,j,t}$.

It is clear that the history of $\{u_{j,t} - \overline{u}_{q+1,j,t}\}$ contains information that would lower the variance of $\overline{\eta}_{q,j,t} + u_{j,t} - \overline{u}_{q,j,t}$ since there are non-zero covariances between $\overline{\eta}_{q,j,t}$ and $u_{j,t-i} - \overline{u}_{q+1,j,t}$. The first of these covariances, for example, will be $(\gamma_1 - \theta_{q,1})\frac{\sigma_u^2}{q(q+1)}$ which will generally be non-zero.

In contrast, the history of $u_{j,t} - \overline{u}_{q,j,t}$ contains no information that would lower the variance of $\overline{\eta}_{q,j,t} + u_{j,t} - \overline{u}_{q,j,t}$ since the covariances between $\overline{\eta}_{q+1,j,t}$ and $u_{j,t-i} - \overline{u}_{q,j,t-i}$ are all zero. To see this note that the first of these covariances will be:

$$\text{covar}((\gamma_1 - \theta_{q+1,1})\overline{u}_{q+1,j,t-1}, (\frac{q-1}{q}u_{j,t-1})$$
$$-\text{covar}((\gamma_1 - \theta_{q+1,1})\overline{u}_{q+1,j,t-1}, (\frac{u_{j+1,t-1} + ... + u_{j+q-1,t-1}}{q})$$

This equals $(\gamma_1 - \theta_{q+1,1})\frac{q-1}{(q+1)q}\sigma_u^2 - (\gamma_1 - \theta_{q+1,1})\frac{q-1}{(q+1)q}\sigma_u^2 = 0$. It follows from this that the other covariances are also zero. Hence the information contained in the history of $\{u_{j,t} - \overline{u}_{q+1,j,t}\}$ would allow the agent to reduce the error variance of a forecast made from information set q, but the information contained in the history of $\{u_{j,t} - \overline{u}_{q,j,t}\}$ would not allow the agent to reduce the error variance of a forecast made from information set $q + 1$. Since the agent with information set $q+1$ can always obtain the series $\eta_{q,j,t}$ from the histories of $\{D_{j,t}....D_{j+j+q,t}\}$ it follows that the agent with information set $q + 1$ has a superior information set which can always be used to produce a forecast with a lower error-variance than the forecast made by an agent with information set q.

Table 1

Forecast Error Variance and the Size of the Information Set, Y

Panel A $\sigma_\varepsilon^2 = 1$

Model : $D_{j,t} = \varepsilon_t + 0.9\varepsilon_{t-1} + u_{j,t} - 0.8u_{j,t-1}$

$Y \mid \sigma_u^2$	0.5	1	2	10
1	2.5312	3.4471	4.9918	14.8051
2	2.2215	3.0312	4.4471	13.9344
5	1.9363	2.6405	3.9211	12.9918
10	1.7901	2.4363	3.6405	12.4471
100	1.5579	2.0988	3.1612	11.4363
∞	1.5000	2.0000	3.0000	11.0000

Panel B $\sigma_\varepsilon^2 = 1$

Model : $D_{j,t} = \varepsilon_t + 0.9\varepsilon_{t-1} + u_{j,t} + 0.5u_{j,t-1}$

$Y \mid \sigma_u^2$	0.5	1	2	10
1	1.6173	2.1472	3.1719	11.2027
2	1.5863	2.1173	3.1472	11.1935
5	1.5507	2.0768	3.1072	11.1719
10	1.5310	2.0507	3.0768	11.1472
100	1.5041	2.0078	3.0146	11.0507
∞	1.5000	2.0000	3.0000	11.0000

References

Akerlof, George A. and Janet L. Yellen (1985) 'Can small deviations from rationality make significant differences to economic equilibria?' *American Economic Review* 75, 708-720.

Ball, Laurence (2000) 'Near-rationality and inflation in two monetary regimes', NBER Working Paper 7988

Buiter, **Willem H.** (1980) 'The Economics of Dr. Pangloss', *Economic Journal*, 90, 34-50.

Conlisk, John (1996) 'Why Bounded Rationality?', *Journal of Economic Literature*, 34(2), 669-700.

Cutler, David M., James M. Poterba and Lawrence H. Summers (1991) 'Speculative Dynamics'. *Review of Economic Studies*, 58, 529-546.

Demery, David and Nigel W. Duck (1999) 'Imperfect information and consumption in the United States and the United Kingdom', *Economica*, 66, 375-387.

Demery, David and Nigel W. Duck (2000) 'Incomplete information and the time series behaviour of consumption', *Journal of Applied Econometrics*, 15, 355-366.

Demery, David and Nigel W. Duck (2001) 'The New Keynesian Phillips Curve and imperfect information', *University of Bristol Working Paper* 01/528

Feige, Edgar and Douglas K. Pearce (1976) 'Economically rational expectations: are innovations in the rate of inflation independent of innovations in measures of monetary and fiscal policy?', *Journal of Political Economy*, 84, 499-522.

Galbraith, John W. (1988) 'Modelling expectations formation with measurement errors' *The Economic Journal*, 98, 412-428.

Galí, Jordi and Gertler, Mark (1999) 'Inflation dynamics: a structural econometric analysis' *Journal of Monetary Economics,* 44, 195-222

Galí, Jordi, Gertler, Mark and López-Salido, J. David (2001) 'European inflation dynamics', *European Economic Review,* 45, 1237-1270.

Goodfriend, Marvin (1992) 'Information-aggregation bias', *American Economic Review*, 82, 508-519.

Hamilton, James D. (1994) *Time Series Analysis.* Princeton University Press, Princeton.

Mankiw, N. Gregory (2001) 'The inexorable and mysterious trade-off between inflation and unemployment', *Economic Journal,* 111, C45-C61.

Mankiw, N. Gregory and Reis, Ricardo (2001) 'Sticky information versus sticky prices: a proposal to replace the New Keynesian Phillips Curve', *Harvard University Mimeo.*

Pesaran (1987) *The Limits to Rational Expectations.* Basil Blackwell Inc. New York.

Pischke, Jörn-Steffen (1995) 'Individual income, incomplete information, and aggregate consumption', *Econometrica,* 63, 805-840.

Woodford, Michael (2001) 'Imperfect common knowledge and the effects of monetary policy' *Mimeo, Princeton University.*