

Notes on Rational Expectations

Gerald P. Dwyer, Jr.
Federal Reserve Bank of Atlanta
e-mail: gdwyer@dwyerecon.com
<http://www.dwyerecon.com>

October 1999

These are brief notes introducing rational expectations and the immediate implications of such expectations. They are not intended to be exhaustive and no doubt are incomplete in places.

I. REVIEW OF EXPECTATIONS AND VARIANCES

Among other things, rational expectations imply that firms' and households' expectations obey the same algebraic operations that apply to mathematical expectations. As a result, reviewing the definitions and implications of expectations and variances will make the implications of rational expectations clearer.

A. Expected values and variances

For a discrete random variable, the expected value of a random variable Y is $E(Y) = \sum_i p_i y_i$. I will suppress the subscript i when I think that no confusion will result.

For a continuous random variable, the expected value of a random variable Y is $E(Y) = \int_y yp(y)dy$.

Expectation is a linear operator. This means that, if a and b are constants,

$$E(a + bY) = a + bE(Y)$$

and

$$E(X + Y) = E(X) + E(Y)$$

A property of expectations that is odd to mention in this context but has more interesting implications in other contexts is $E[E(Y)] = E(Y)$. This means the following. Suppose that you have the expected value of Y , namely $E(Y)$. Taking the expectation of this with respect to the same probability distribution function (pdf) results in $E(Y)$. This is not surprising — $E(Y)$ is a constant relative to the pdf and a constant goes through the expectations operator, whether it is denoted a or $E(Y)$.

Variance is defined as $E[Y - E(Y)]^2 = \sum_i p_i [y_i - E(Y)]^2$ and analogously for continuous random variables. Variance is not a linear operator:

$$\text{Var}(a + bY) = b^2\text{Var}(Y)$$

and

$$\text{Var}(X + Y) = \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y)$$

where

$$\text{Cov}(X, Y) = E[X - E(X)][Y - E(Y)].$$

B. Conditional Expected Values and Variances

The expected value of Y conditional on X is $E(Y|X = x) = \sum_i y_i p(y_i|X = x)$, where the conditional probability $p(y_i|X = x)$ is the probability of a value y_i conditional on a particular value x of the random variable X. Let $p(x, y)$ be the joint probability of the values x and y and let $p(x)$ be the marginal probability: the probability of the value x for all possible values of Y, i.e. $p(x) = \sum_i p(x, y_i)$. Then the conditional probability is defined implicitly by $p(x, y) = p(y|X = x)p(x)$ and $p(y|X = x) = p(x, y) / p(x)$ if $p(x) \neq 0$.

1. Pertinent properties of conditional expectations

Pertinent properties of conditional expectations can be verified relatively easily. Among other things, the expected value of Y can be recovered from the conditional expected value of Y by weighting the conditional expected values by their probabilities for all possible values of X, i.e.,

$$E(Y) = \sum_i E[Y|X = x_i]p_i.$$

If X and Y are independent

$$E[Y|X = x] = E(Y).$$

For a continuous function $g(\cdot)$ that is not linear,

$$E[g(X, Y)|X = x] = E[g(x, Y)|X = x]$$

but

$$E[g(X, Y)|X = x] \neq g(E[X|X = x], Y).$$

The last equation says that the conditional expected value of a nonlinear function is not equal to the nonlinear function of the conditional expected value. (This also holds for unconditional expectations

as well.) The inability to get the expected value of a nonlinear function by using expected values in the equation underlies the common use of linear equations in rational expectations models.

2. Joint normality

If two variables are jointly normally distributed then

$$E[Y|X = x] = EY + \frac{\text{Cov}(X, Y)}{\text{Var}(X)}[x - E(X)].$$

You should recognize this as the linear least squares formula. However, this equation is not an estimating equation: it holds in the population. This relationship between least squares and linearity of the conditional expected value underlies the common assumption of normality in the theoretical literature using rational expectations. While it is possible to look at only linear relationships even if variables are not jointly normal, the resulting relationship is not the same as the conditional expectation.

II. RATIONAL EXPECTATIONS

What is the definition of rational expectations and what are the implications of such expectations?

I will assume normality throughout this discussion. This does not mean that variables cannot be serially correlated. (The more precise usage would be to say that I assume Gaussianity — a Gaussian distribution has serially uncorrelated innovations but can have other linear structure. If this elaboration seems incomprehensible or not helpful, don't worry about it.) While I could dispense with it, assuming normality vastly simplifies the discussion and dispensing with normality would add little.

Rational expectations can be defined in a variety of ways.

A simple way of thinking about rational expectations is that, given the information available to agents (i.e. firms, households, the government and whoever and whatever else might appear in an economy), they cannot predict their own forecast errors. Let Y^e denote the agent's anticipated value of Y . This says that $(Y - Y^e)^e$, where the superscript e denotes the anticipation operator. It seems difficult to think that anticipations could violate such a condition unless someone were schizophrenic. While seemingly innocuous, it is not really. Effectively, it is a consistency criterion on the anticipation operator because it requires that $Y^e - (Y^e)^e = 0$.

A related definition of rational expectations would define rational expectations as being expectations that use available information in a way that the agent could not use the information available to make himself or herself better off. This is more stringent, in part requiring that people

use the information available to make the best forecast that they can, where *best* is defined in terms of the agent's objective function and ability to make forecasts.

A more restrictive definition requires that rational expectations be the same as the predictions of the relevant economic theory. This is Muth's definition. Among other things, it requires that a theory not imply that agents' forecast errors are predictable given information available to agents and the relevant economic theory.

An even more restrictive definition requires that the subjective and objective parameters be the same — e.g. the covariance between X and Y that agents use to make forecasts is the same as actual covariance. This definition is due to Lucas and Prescott. While wildly unrealistic, this definition can be quite useful and in some circumstances does a good job of characterizing people's behavior.

Given joint normality and the identification of agents' forecasts with mathematical expectations, certain propositions follow.

1. Forecast error unpredictable. Denote the rational expectation by $E[y_{t+1}|I_t]$, where y_{t+1} is the variable in period $t+1$ and I_t is the information available in period t to forecast y . This is a conditional expectation with more economical notation. Let the forecast error be $\varepsilon_{t+1} = y_{t+1} - E[y_{t+1}|I_t]$. This just says that $E[\varepsilon_{t+1}|I_t] = 0$. This holds because $E[\varepsilon_{t+1}|I_t] = E\{[y_{t+1} - E[y_{t+1}|I_t]]|I_t\} = E[y_{t+1}|I_t] - E[y_{t+1}|I_t] = 0$.

2. Law of iterated expectations. Suppose that there are two information sets that can be used to predict y and that the information set I is a subset of the information set Ω . Start with the expected value of y conditional on Ω , $E[y_{t+1}|\Omega_t]$. Then take the expectation conditional on I of that expectation. The law of iterated expectations implies that $E\{E[y_{t+1}|\Omega_t]|I_t\} = E[y_{t+1}|I_t]$.

The usefulness of this proposition is not immediately obvious but it often is crucial for doing tests. A theory often implies something about the expected value conditional on all the information available to agents, $E[y_{t+1}|\Omega_t]$, but an econometrician only observes the outcome y_{t+1} and a *subset* of the information available to agents I_{t+1} . This proposition implies that the behavior of the expected value of y_{t+1} conditional on the information available to the econometrician, namely $E[y_{t+1}|I_t]$, is related in a particular way to agents' information conditional on their information Ω_t . Knowing this, it often is possible to construct tests even though there may be no way of knowing all of the contents of the information set available to agents. More on this later in the class.

3. One-step-ahead forecast errors are serially uncorrelated if agents' know their own forecast errors in all past periods. Let f_{t+1} denote the forecast made at t for period $t+1$. Suppose that the variable y_{t+1} is being forecasted. Let $\eta_{t+1} = y_{t+1} - f_{t+1}$. If expectations are rational, then $f_{t+1} = E[y_{t+1}|\Omega_t]$. We know that

$$E[\eta_{t+1}|I_t] = E[y_{t+1}|I_t] - E[f_{t+1}|I_t] = E[y_{t+1}|I_t] - E[E[y_{t+1}|\Omega_t]|I_t] = 0.$$

Suppose that the information in I_t includes past forecast errors $\eta_t, \eta_{t-1}, \dots$. Then the equation above and normality implies that a regression of the current forecast error on the forecast error in $t, t-1$, etc., would result in a set of coefficients equal to zero. (Why?) A regression of η_{t+1} on $\eta_t, \eta_{t-1}, \dots$ having coefficients equal to zero is equivalent to saying that the forecast errors are serially uncorrelated. (Unless you've had a time series course, this equivalence may not be completely obvious and you'll have to take my word for it.)

4. Chain rule of forecasting. How are forecasts at period t for period $t+k$ made? Let ${}_t f_{t+k}$ be the forecast made at t for $t+k$. Assume rational expectations hold, so that ${}_t f_{t+k} = E[y_{t+k} | \Omega_t]$. Suppose that people do not forget and that they may learn some things as time goes on. This implies that Ω_{t+k-1} is a superset of Ω_{t+k-2} is a superset of Ω_{t+k-3} ... is a superset of Ω_t .

$${}_t f_{t+1} = E[y_{t+1} | \Omega_t]$$

$$\text{Then } {}_t f_{t+2} = E[y_{t+2} | \Omega_t] = E\{E[y_{t+2} | \Omega_{t+1}] | \Omega_t\}$$

$${}_t f_{t+3} = E[y_{t+3} | \Omega_t] = E\{E\{E[y_{t+3} | \Omega_{t+2}] | \Omega_{t+1}\} | \Omega_t\}$$

and so on. The name "chain rule of forecasting" comes from the fact that, to get $E[y_{t+3} | \Omega_t]$, one first computes $E[y_{t+3} | \Omega_{t+2}]$. Then one computes the expected value of this conditional on Ω_{t+1} using the law of iterated expectations, i.e., $E[y_{t+3} | \Omega_{t+1}] = E\{E[y_{t+3} | \Omega_{t+2}] | \Omega_{t+1}\}$. Then one computes $E[y_{t+3} | \Omega_t] = E\{E\{E[y_{t+3} | \Omega_{t+2}] | \Omega_{t+1}\} | \Omega_t\}$. This forms a chain of computations in sequence that eventually lead to the answer.

An example using the chain rule of forecasting may make this statement less abstract. Suppose that the variable y is a first-order autoregression

$$y_t = \beta y_{t-1} + \varepsilon_t$$

with no constant, β is the slope parameter and ε_t is the innovation with

$$E\varepsilon_t = 0 \quad \text{and} \quad E\varepsilon_t \varepsilon_s = \begin{cases} \sigma^2, & t = s \\ 0, & t \neq s \end{cases}$$

Notice that the covariance of the innovations is zero. Suppose that it is desired to compute a three-step-ahead forecast, i.e., compute $E[y_{t+3} | y_t]$. It may seem desirable to consider past values of y other than just y_t but we need not do so for a first-order autoregression, as the following computations show. The notation below follows common practice and identifies the sequence $\{y_t\}$ with the information set. The information set Ω_t is identified by y_t , Ω_{t+1} is identified by y_{t+1} , and so on.

How can we get a forecast for period $t+3$ in period t ? A one-step-ahead forecast at period t based on y_t is

$$E[y_{t+1}|y_t] = \beta y_t.$$

The use of period t is not crucial here and we can just as well write

$$E[y_{t+3}|y_{t+2}] = \beta y_{t+2}$$

for a forecast at period $t+2$ for period $t+3$. At period t , this does not deliver what we want because we do not know y_{t+2} . In period t , we do not know y_{t+2} but we do know that

$$E[y_{t+2}|y_{t+1}] = \beta y_{t+1}.$$

This gives us a chain relating a forecast three periods from now to one-step-ahead forecasts. The chain is

$$\begin{aligned} E[y_{t+1}|y_t] &= \beta y_t \\ E[y_{t+2}|y_{t+1}] &= \beta y_{t+1} \\ E[y_{t+3}|y_{t+2}] &= \beta y_{t+2} \end{aligned}$$

This is not quite enough though because it includes the unknown values of y_{t+1} and y_{t+2} . We can compute the values of $E[y_{t+2}|y_t]$ and $E[y_{t+3}|y_t]$ by using the law of iterated expectations.

How can we compute the value of $E[y_{t+2}|y_t]$? We have that

$$E[y_{t+2}|y_{t+1}] = \beta y_{t+1}.$$

Taking expectations of both sides conditional on y_t yields

$$E[y_{t+2}|y_t] = \beta E[y_{t+1}|y_t].$$

This is what we want to be able to compute $E[y_{t+2}|y_t]$. In terms of the first-order autoregression, we know that $E[y_{t+1}|y_t] = \beta y_t$. Hence,

$$E[y_{t+2}|y_t] = \beta E[y_{t+1}|y_t] = \beta[\beta y_t] = \beta^2 y_t.$$

The expected value of y_{t+3} can be found similarly. We have that

$$E[y_{t+3}|y_{t+2}] = \beta y_{t+2}$$

Taking expectations of both sides of this equation conditional on y_{t+1} yields

$$E[y_{t+3}|y_{t+1}] = \beta E[y_{t+2}|y_{t+1}].$$

Because y_{t+1} is no more known than y_{t+2} , we want to take expectations conditional on y_t .

$$E[y_{t+3}|y_t] = \beta E\{E[y_{t+2}|y_{t+1}]|y_t\}.$$

We know that $E[y_{t+2}|y_{t+1}] = \beta y_{t+1}$ and that $E[y_{t+1}|y_t] = \beta y_t$. By the law of iterated expectations,

$$E[y_{t+3}|y_t] = \beta E[\beta y_{t+1}|y_t] = \beta^2 E[y_{t+1}|y_t] = \beta^2 [\beta y_t] = \beta^3 y_t.$$

In short, just use repeated substitution to get the implied forecast.

If one is computing a sequence of forecasts for periods t , $t+1$, $t+2$, ..., $t+k$, then it only is necessary to use the predicted value for the prior period. For example, $E[y_{t+3}|y_t] = \beta E[y_{t+2}|y_t]$. Once it is known that $E[y_{t+2}|y_{t+1}] = \beta^2 y_t$, this can be substituted in the equation above. The chain becomes

$$\begin{aligned} E[y_{t+1}|y_t] &= \beta y_t \\ E[y_{t+2}|y_t] &= \beta E[y_{t+1}|y_t] \\ E[y_{t+3}|y_t] &= \beta E[y_{t+2}|y_t] \end{aligned}$$

This algebraic statement is for a first-order autoregression, but the basic implication is just this: use forecasts of future values conditioned on prior values to eliminate expected values of future values from a forecasted value.