Extensions to Notes on Macroeconomics

Thomas J. Sargent

July 1976

Working Paper #: 64

Rsch. File #: 299.1

Not for Distribution

Extensions to Notes on Macroeconomics

by

Thomas Sargent

July 1976

Contents

In Defense of Keynesian Analyses That "Ignore" the Government's Budget Constraint

Following the appearance of Carl Christ's article, it has been common to hear the textbook Keynesian model criticized for ignoring the fact that the government has a budget constraint. Christ and others have claimed that various Keynesian multiplier formulas are mistaken because they ignore the government's budget constraint. The assertion is that the government's budget constraint plays a role which Keynesian economists overlooked, and which alters some of the substantive conclusions of Keynesian models. Here I argue that the textbook Keynesian model is totally immune from this charge, and that the implications that Christ has drawn from the existence of the government budget constraint are at best innocuous and at worst false.

This note is in the nature of a defense of old fashioned macroeconomic analysis and obviously makes no pretense at being original. Some time ago James Tobin presented a very clear and correct nontechnical statement of the role of the government's budget constraint in Keynesian analysis in his "Deficit, Deficit, Who's Got the Deficit?". Neil Wallace showed formally how ordinary static Keynesian analyses are immune to the criticisms made by Christ and others. Despite these contributions, at conferences and in print the charge continues to be made that Keynesian analyses are negligent in neglecting the government's budget constraint.

Consider the following standard Keynesian macroeconomic model:

1. $Y = F(N, K)$, $F_N, F_K, F_{NK} > 0$; $F_{KK}, F_{NN} < 0$ Production function

2. $w/p = F_N(N, K)$ marginal equality for labor

3. $I = I(F_K(N, K) - (r - \pi))$ $I' > 0$, investment schedule

4. $C = C(Y - T)$ $1 > C' > 0$ consumption schedule

5. $C + I + G = Y$ national income identity

6. $M/p = m(r, Y)$ $m_r < 0, m_Y > 0$ portfolio balance condition

7. $\frac{\dot{w}}{w} = f(\frac{N^s - N}{N^s}) + \alpha\pi$ $f' < 0, 0 \leq \alpha \leq 1.$ Phillips curve

8. $\dot{K} = I$

9. $\dot{M} + \dot{B} = p(G - T)$

Here Y is real GNP, N employment, K the capital stock, w the money wage, p the price level, I real investment, C real consumption, G is real government purchases, T is real tax collections, M is the stock of money (a government liability), r is the interest rate, π is the expected rate of inflation, N^s is the labor supply, and B is the stock of government bonds held by the public. The variables Y, C, I, G, and T are all real flows measured in goods per unit time; M and B are stocks measured in dollars; N, N^s, and K are stocks; π and r are pure numbers per unit time. Each of the variables appearing in the system of equations (1) - (9) should be regarded as a function of time. We will assume that time passes continuously.[2]

The model is usually manipulated like this.[3] One feeds into the model paths of the exogenous variable $G(t)$, $T(t)$, $M(t)$, $N^s(t)$, and

$\pi(t)$ over the time interval $[t_0, t_1)$, $t_1 > t_0$, where t_1 may be infinity. In addition, one must specify initial conditions for K, w, and B in the form of their values at time t_0, namely the single values $K(t_0)$, $w(t_0)$, and $B(t_0)$. Given these inputs, the model consisting of equations (1) - (9) will generate entire time paths over the interval $[t_0, t_1)$ for the endogenous variables Y, N, p, r, C, and I, and paths over the interval (t_0, t_1) for the remaining three endogenous variables K, w, and B.

Two kinds of analysis are commonly carried out with this model. Comparative dynamics examines the alternative time paths of the endogenous variables associated with alternative time paths of the exogenous variables. Comparative statics freezes time and asks how the endogenous variables at one single point in time would differ in response to differing assumed values of the exogenous variables at that point in time. It is important to note that some variables that are endogenous from the dynamic point of view, in particular K, w and M+B in this model, are exogenous from a static point of view, in the sense that they are inherited from the past and cannot jump at a point in time in order to equilibrate the model in response to assumed jump discontinuities in the exogenous variables at a point in time. That is, w, K and M+B are determined by the solutions of the differential equations (7), (8), and (9) and are therefore continuous functions of time which are given from the point of view of a static or point in time analysis. On the other hand, the variables Y, N, p, r, C and I are free to exhibit jump discontinuities in order to equilibrate the model. Consequently, these six variables are endogenous both from the dynamic and static points of view.

Now consider the following two propositions:

Proposition 1: The time path of M matters in the sense of affecting some of the endogenous variables Y, N, p, r, C, and I.

Proposition 2: For determining the paths of Y, N, p, r, C and I it is important to take into account equation (9), the government budget constraint.

Proposition 1 has long been acknowledged in competent presentations of the Keynesian model.[4] Christ and some subsequent writers have advanced proposition 2 as a truth seemingly over and above Proposition 1 , and have criticized the Keynesian analysis for failing to take it into account. However, at best proposition (2) is a restatement of part of proposition 1. At worst, it is just plain wrong.

First, notice that from a static point of view, equations 1 - 6 form a subset of equations that determines Y , N, p, r, C, and I at a point in time, given the exogenous variables K, M, G, T and π at that point in time. In particular, notice that Y, N, p, r, C, and I are determined prior to and independently of $\dot{M} + \dot{B}$, which is determined by (9). From a static point of view, then, it is perfectly legitimate to ignore (9) in determining Y, N, p, r, C, and I because $(\dot{M} + \dot{B})$ does not appear in equations (1) - (6).

So proposition 2, if it is saying anything, must be about dynamics. But it seems wrong even as a statement about dynamics. This unfavorable interpretation of proposition 2 is obtained by noticing that our model is dynamically recursive with respect to B. Equation (9), the government budget constraint, determines only the path of $B(t)$, for t on the interval (t_0, t_1). Removal of equation (9) from the system, as is often done in textbook analyses of Keynesian models, in no way limits the ability of the model to determine time paths of the "interesting" endogenous variables Y, N, p, r, C and I.

The most favorable interpretation that can be put on proposition (2) is as follows. Suppose that one wants to manipulate the model under a regime[5] in which $B(t) = B(t_0)$ for all t in $[t_0, t_1)$. Under this regime, all that the model user is free to specify exogenously about the path of M is the initial condition $M(t_0)$. Under the constraint $B(t) = B(t_0)$, equation (9) will determine $\dot{M}$, and ultimately an entire time path of M on the open interval (t_0, t_1). The time path of $M(t)$ so determined will be continuous, by virtue of being the solution of the differential equation (9). Notice that this regime rules out open market operations in the form of jump discontinuities in M that leave $M + B$ a continuous function of time. In this special regime, Proposition 2 is true, because things have been set up so that equation (9) determines the time path of M ---something that it generally does not do in a regime which permits open market operations. But notice that Proposition 2 adds absolutely nothing to what has already been asserted in Proposition (1). So in this special case, Proposition (2) is correct but innocuous. It is not correct to criticize users of the Keynesian model, who readily grant proposition 1, for failing to mention Proposition 2.

One way that might seem to salvage some meaning to proposition (2) is to amend the system by replacing (4) and (6) with

$$(4') \qquad C = c(Y - T - \frac{M + B}{p}\pi) \qquad 1 > c' > 0$$

$$(6') \qquad \frac{M}{p} = m(r, Y, \frac{M + B}{p} + K) \qquad m_r < 0, \quad m_Y > 0$$

$$1 > m_{\frac{M+B}{p} + K} > 0 \quad .$$

Here we have entered the real value of government debt to the public in the definition of disposable income and in the portfolio balance schedule in the usual ways. In the model formed by replacing (4) and (6) with (4') and (6'), not only the division of government debt between M and B, but also the total stock of government debt $M + B$ matters in the sense of affecting the variables Y, N, p, r, C, and I. Since equation (9) determines the path of $M + B$ over time, it is important to take it into account. This is true, but is subsumed in the following version of Proposition (1), which is appropriate for a system in which equations like (4') or (6') replace (4) or (6):

Proposition 1' : The time paths of M and $M + B$ matter in the sense of affecting some of the endogenous variables Y, N, p, r, C and I.

This proposition was long accepted in presentations of the Keynesian model, as illustrated in the extensive literature on the

Pigou effect and the burden of the debt.[6] Given the acceptance of Proposition 1', Proposition 2 adds nothing. While the system with (4') and (6') put Proposition 2 in the most favorable possible light, it is not the system that Christ used in his attack on the logic of Keynesian multiplier formulas.[7]

Footnotes

1 For example, see Ackley and Bailey.

2 The exogenous variables M, G, T, π, and N^s are assumed to be right continuous functions of time which possess right hand time derivatives everywhere. A dot over a variable denotes a right hand time derivative. Notice that a variable, say $M(t)$, can have a discontinuity from the left at t and still possess a right hand time derivative.

3 This is the standard regime to impose on the model, but is not the only possible one. By a regime I mean a categorization of variables into exogenous and endogenous classes.

4 For example, see Ackley and Bailey.

5 The other exogenous variables continue to be time paths of G, T, and π over $[t_0, t_1)$ and initial conditions $K(t_0)$ and $w(t_0)$.

6 For example, see Metzler.

7 Christ has criticized the static Keynesian multiplier formulas because they do not agree with the long run multipliers he obtains by imposing some (in my opinion, strange) stationarity conditions--- in particular, the requirement that in the long run the government deficit be zero. The Keynesian static multiplier formulas, which pertain to point-in-time exercises, are simply not commensurate

with the long run dynamic multipliers that Christ calculates. In addition, notice that Christ's long run multipliers ignore the process of capital accumulation and price and wage dynamics, and so don't correspond to a solution of the full Keynesian model.

References

Ackley, Gardner, Macroeconomic Theory, New York, Macmillan, 1961.

Bailey, Martin J., National Income and the Price Level, New York, McGraw-Hill, 1962.

Christ, Carl F., "A Simple Macroeconomic Model with a Government Budget Restraint," Journal of Political Economy, Vol. 76, Jan.-Feb. 1968.

Metzler, Lloyd, "Wealth, Saving and the Rate of Interest," Journal of Political Economy, 1951.

Tobin, James, National Economic Policy, New Haven, Yale University Press, 1966.

Wallace, Neil, "A Static Nonstationary Analysis of the Interaction Between Monetary and Fiscal Policy," University of Minnesota, Center for Economic Research, Discussion Paper No. 9, August 1971.

Linear Least Squares Projections (Regressions)

The concept of linear regression has many important uses in macroeconomics, several of which we shall illustrate in subsequent chapters. One very important application will be its use in modeling the "signal extraction" problem faced by agents in an environment in which they have imperfect information about a variable affecting their welfare. By using a linear regression, agents can estimate that unobserved variable in a manner that is optimal, in a certain sense. Two leading applications of the signal extraction model in macroeconomics are Robert Lucas's model of the Phillips curve and Milton Friedman's theory of the consumption function. Another use we will make of linear regression is to characterize and study the optimal control problem facing the monetary and fiscal authorities.

Linear Least Square Regression: The Orthogonality Condition

We consider a set of random variables $y, x_1, x_2, \ldots, x_n$. The population means of this list of random variables are denoted $Ey, Ex_1, \ldots, Ex_n$. We assume that these means are finite, as are the population second moments $Ey^2, Ex_1^2, \ldots, Ex_n^2$. By the Cauchy-Schwarz inequality the following cross-second moments exist and are finite:

$$\begin{array}{llll} Eyx_1 & Ex_1^2 & Ex_1x_2 \ldots & Ex_1x_n \\ Eyx_2 & Ex_2x_1 & Ex_2^2 \quad \ldots & Ex_2x_n \\ \vdots & \vdots & & \\ Eyx_n & Ex_nx_1 & Ex_nx_2 \ldots & Ex_n^2 \end{array}$$

Consider estimating the random variable y on the basis of knowing values only for the random variables $x_1, \ldots, x_n$ as well as knowing all of the means and second moments listed above.[1/] More specifically, suppose we restrict ourselves to estimating y by the linear function of the x_i's,[2/]

(1) $$\hat{y} = a_0 + a_1x_1 + \ldots + a_nx_n.$$

We seek to choose the a_i's so that the random variable $\hat{y}$ is as "close" to y as possible, in the least squares sense that $E(y-\hat{y})^2$ is a minimum. Thus, our problem is to minimize

(2) $$E(y-(a_0+a_1x_1+\ldots+a_nx_n))^2$$

with respect to $a_0, a_1, \ldots, a_n$. To facilitate the computations, let us define the new (trivial) random variable $x_0 \equiv 1$.

We are now in a position to state the orthogonality principle: A necessary and sufficient set of conditions for $a_0, a_1, \ldots, a_n$ to minimize (2) is

---

1/ The reader with some background in econometrics will note that we are <u>not</u> studying the "general linear model," (e.g., see J. Johnston, <u>Econometric Methods</u>, Chapter 4), which assumes that the right-hand side x variables are nonstochastic.

2/ The restriction to a linear function is in general a binding one. It is possible to show that to minimize $E\{y-g(x_1, \ldots, x_n)\}^2$ with respect to the choice of $g(x_1, \ldots, x_n)$, the optimal thing to do is to set $g(x_1, \ldots, x_n) = E[y|x_1, \ldots, x_n]$, the mathematical expectation of y conditional on $x_1, \ldots, x_n$. In general, the mathematical expectation $E[y|x_1, \ldots, x_n]$ is <u>not</u> a linear function of $x_1, \ldots, x_n$. In the special case that the variates $(y, x_1, \ldots, x_n)$ follow a multivariate normal distribution, the conditional mathematical expectation $E[y|x_1, \ldots, x_n]$ is linear in $x_1, \ldots, x_n$.

$$(3) \quad E\{(y-(a_0+a_1x_1+\ldots+a_nx_n))x_i\} = 0, \qquad i=0, 1, \ldots, n.$$

Condition (3) says that $E(y-\hat{y})\cdot x_i = 0$ for all i. Two random variables w and z are said to be orthogonal if $E(w\cdot z) = 0$. Thus, (3) asserts that $(y-\hat{y})$ is orthogonal to each x_i, i=0, ..., n. The orthogonality principle asserts that the condition $E(y-\hat{y})x_i = 0$ for each i uniquely determines $\hat{y}$. (It will also uniquely determine the a_i's if there is no linear dependence among the x_i's.)

To prove the orthogonality principle, suppose that the a_i's satisfy (3). Consider any other linear estimator of y, say $A_0 + A_1x_1 + \ldots + A_nx_n$, where the A_i's are fixed numbers. The mean squared error associated with using $\sum_{i=0}^{n} A_ix_i$ to estimate y is

$$(4) \qquad E(y-\sum_{i=0}^{n} A_ix_i)^2 = E((y-\sum_{i=0}^{n} a_ix_i)+\sum_{i=0}^{n}(a_i-A_i)x_i))^2$$

$$= E(y-\sum_{i=0}^{n} a_ix_i)^2 + E(\sum_{j=0}^{n}(a_j-A_j)x_j)^2$$

$$+ 2E(y-\sum_{i=0}^{n} a_ix_i)(\sum_{j=0}^{n}(a_j-A_j)x_j)$$

Since the (a_j-A_j)'s are constants, the last term can be rewritten

$$2\sum_{j=0}^{n}(a_j-A_j)E(y-\sum_{i=0}^{n} a_ix_i)\cdot x_j = 0,$$

since the orthogonality condition (3) is in force, implying $E(y-\sum_{i=0}^{n} a_ix_i)x_j = 0$ for all j. Therefore, (4) becomes

$$E(y-\sum_{i=0}^{n} A_ix_i)^2 = E(y-\sum_{i=0}^{n} a_ix_i)^2 + E(\sum_{j=0}^{n}(a_j-A_j)x_j)^2$$

$$\geq E(y-\sum_{i=0}^{n} a_ix_i)^2,$$

with equality holding for $a_j=A_j$. The orthogonality condition is therefore a sufficient condition for minimizing the mean squared error (2). (It can also be shown to be a necessary condition.)

The orthogonality condition (3) in effect asserts that the "forecast error" $y-\sum_{i=0}^{n} a_i x_i$ is orthogonal to each of the x_i's and therefore is also orthogonal to any linear combination of the x_i's. Defining the forecast error as $\varepsilon = y-\sum_{i=0}^{n} a_i x_i$, we therefore have

$$(5) \qquad y = \sum_{i=0}^{n} a_i x_i + \varepsilon$$

where $E(\varepsilon \cdot \sum_{i=0}^{n} a_i x_i) = 0$

and $E\varepsilon x_i = 0$ for i=0, 1, ..., n.

Thus, (5) decomposes y into orthogonal parts. By virtue of the orthogonality of the random variables $\sum_{i=0}^{n} a_i x_i$ and ε, we have the decomposition

$$Ey^2 = E(\sum_{i=0}^{n} a_i x_i)^2 + E\varepsilon^2$$

The random variable $\sum_{i=0}^{n} a_i x_i$, where the a_i's are chosen to satisfy the least squares orthogonality condition (3), is called the <u>projection</u> of y on $x_0, x_1, \ldots, x_n$. We will find it convenient to denote the projection of y on $x_0, x_1, \ldots, x_n$ as

$$\sum_{i=0}^{n} a_i x_i \equiv P[y|1, x_1, x_2, \ldots, x_n]$$

where remember that $x_0 = 1$ identically.

The orthogonality conditions (3) can be readily rearranged in the form of the familiar least squares normal equations. Write (3) explicitly for i=0, 1, ..., n to get the normal equations

$$
(6) \quad \begin{bmatrix} Ey \\ Eyx_1 \\ Eyx_2 \\ \vdots \\ Eyx_n \end{bmatrix} = \begin{bmatrix} 1 & Ex_1 & Ex_2 & \cdots & Ex_n \\ Ex_1 & Ex_1^2 & Ex_1x_2 & \cdots & Ex_1x_n \\ Ex_2 & Ex_1x_2 & Ex_2^2 & & \\ \vdots & & & & \\ Ex_n & Ex_1x_n & Ex_2x_n & \cdots & Ex_n^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} .
$$

Assuming that the (n+1)x(n+1) matrix above has an inverse, we have the following explicit equation for the a_i's:

$$
(7) \quad \begin{bmatrix} a_0 \\ \alpha_1 \\ \vdots \\ a_n \end{bmatrix} = [Ex_ix_j]^{-1}[Eyx_k]
$$

where $[Ex_ix_j]^{-1}$ is the inverse of the matrix with i+1, j+1st element Ex_ix_j, and $[Eyx_k]$ is the (n+1)x1 vector with k+1th element Eyx_k.

As an example, consider projecting y against a single variate x_1 (as well as the trivial variate $x_0=1$). Then (6) becomes

$$
\begin{bmatrix} Ey \\ Eyx_1 \end{bmatrix} = \begin{bmatrix} 1 & Ex_1 \\ Ex_1 & Ex_1^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} .
$$

The solution of these two equations turns out to be

$$
a_1 = \frac{E(y-Ey)(x_1-Ex_1)}{E(x_1-Ex_1)^2}
$$

$$
a_0 = Ey - a_1Ex_1 \quad .
$$

Denote the covariance between y and x_1 as $\sigma x_1 y = E\{(y-Ey)(x_1-Ex_1)\}$ and the variance of x_1 as $\sigma^2 x_1 = E\{(x_1-Ex_1)^2\}$. Then the equations for a_1 and a_0 become the familiar

$$(8)\qquad \begin{aligned} a_1 &= \frac{\sigma x_1 y}{\sigma x_1^2} \quad . \\ a_0 &= Ey - a_1 Ex_1 \end{aligned}$$

Recursive Projection

It happens that the simple univariate formulas (8) can be used in a recursive way to assemble projections on many variables, e.g., $P[y|1, x_1, \ldots, x_n]$. This often affords a computational saving, and also carries insights about sequential learning.

Write the decomposition (5) for n=2 as

$$y = P[y|1, x_1, x_2] + \varepsilon$$

$$(9)\qquad y = a_0 + a_1 x_1 + a_2 x_2 + \varepsilon$$

where $E\varepsilon=0$, $E\varepsilon x_1=0$, and $E\varepsilon x_2=0$. These three orthogonality conditions insure that the a_i's are the least squares parameter values. Now project both sides of (9) against 1 and x_1 to obtain the equation

$$(10)\qquad P[y|1, x_1] = a_0 + a_1 x_1 + a_2 P[x_2|1, x_1].$$

To get from (9) to (10) we have used the facts that

$$P[a_0|1, x_1] = a_0$$

$$P[x_1|1, x_1] = x_1$$

$$P[\varepsilon|1, x_1] = 0$$

The first two came directly from application of the orthogonality principle to the problem of computing the indicated projection. More directly, it is clear that

$$E\{a_0-t_0-t_1x_1\}^2$$

is minimized by setting $t_0=a_0$ and $t_1=0$. Similarily, $E\{x_1-t_0-t_1x_1\}^2$ is minimized by setting $t_0=0$ and $t_1=1$. The last of the three equalities above comes from noting that from the orthogonality conditions in (9), $E\varepsilon=E\varepsilon x_1=0$. Substituting these into the least squares normal equations for $P[\varepsilon|1, x_1]$ shows that $P[\varepsilon|1, x_1] = 0$.

Subtracting (10) from (9) gives

$$y - P[y|1, x_1] = a_2(x_2-P[x_2|1, x_1]) + \varepsilon \tag{12}$$

where we repeat that $E\varepsilon=E\varepsilon x_1=E\varepsilon x_2=0$. Let $P[x_2|1, x_1] = b_0 + b_1x_1$. The orthogonality conditions imply that

$$E[\varepsilon\cdot(x_2-P[x_2|1, x_1])]$$

$$= E[\varepsilon(x_2-b_0-b_1x_1)]$$

$$= E\varepsilon x_2 - b_0E\varepsilon - b_1E\varepsilon x_1 = 0$$

Thus ε is orthogonal to $x_2 - P[x_2|1, x_1]$. The orthogonality principle therefore implies that $a_2(x_2-P[x_2|1, x_1])$ must be the projection of $y - P[y|1, x_1]$ against $(x_2-P[x_2|1, x_1])$. Thus, (12) can be rewritten

$$y = P[y|1, x_1] + P[(y-P[y|1, x_1])|(x_2-P[x_2|1, x_1])] + \varepsilon \tag{13}$$

Notice that by virtue of the orthogonality conditions on ε, (13) implies

$$P[y|1, x_1, x_2] = P[y|1, x_1] + P[(y-P[y|1, x_1])|(x_2-P[x_2|1, x_1])].$$

Let $P[y|1, x_1] = c_0 + c_1x_1$

$$P[x_2|1, x_1] = b_0 + b_1x_1$$

$$P[(y-P[y|1, x_1])|(x_2-P[x_2|1, x_1])$$
$$= d_0 + d_1(x_2-b_0-b_1x_1).$$

(Actually, from (12), we know that d_0=0.) Then (13) can be written

$$y = c_0 + c_1x_1 + d_0 + d_1(x_2-b_0-b_1x_1) + \varepsilon$$

(14) $$y = c_0 + d_0 + (c_1-b_1d_1)x_1 + d_1x_2 + \varepsilon$$

Comparing (9) with (14), we have

$$a_0 = c_0 + d_0 = c_0$$

(15) $$a_1 = (c_1-b_1d_1)$$

$$a_2 = d_1$$

The relations (15) give the coefficients in the bivariate projection $P[y|1, x_1, x_2]$ in terms of the parameters of three univariate projections.

Equation (13) is a useful description of optimal least squares learning or sequential estimation. If at first we have data only on a variable x_1, the linear least squares estimates of y and x_2 are $P[y|1, x_1]$ and $P[x_2|1, x_1]$, respectively. If an observation x_2 subsequently becomes available, our estimate of y can be improved by adding to $P[y|1, x_1]$ the projection of the unobserved "forecast error"

$y - P[y|1, x_1]$ on the observed forecast error $x_2 - P[x_2|1, x_1]$. So long as these forecast errors are correlated, the new observation on x_2 carries information useful for estimating y.

By induction (or by suitably interpreting x in (13) as a vector of random variables) it is straightforward to extend (13) to the vector form

$$(16) \qquad P[y|\Omega, x] = P[y|\Omega] + P[(y-P[y|\Omega])|(x-P[x|\Omega])]$$

where Ω is a list of random variables. The practical implication of (13) and (16) is that the multivariable regression (16) can be built up sequentially from a set of univariate regressions.

The recursive relation (16) is the foundation of "Kalman filtering," a technique widely used by engineers. We shall see how the sequential learning mechanism in (16) was exploited by Lucas in obtaining his model of the Phillips curve.

The Signal Extraction Problem

Suppose an agent wants to estimate a random variable s but only "sees" the random variable x which is related to s by

$$x = s + n$$

where $Esn=0$; $Es^2, En^2 < \infty$; $Es=En=0$. The linear least squares estimate of s is

$$P[s|1, x] = a_0 + a_1 x$$

The least squares normal equations become

$$a_1 = \frac{E(xs)}{Ex^2} = \frac{E((s+n)s)}{E(s+n)^2}$$

$$a_1 = \frac{Es^2}{Es^2+En^2}$$

$$a_0 = 0.$$

As a slightly richer example, consider a worker who wants to estimate (the log of) his real wage, w-p. He "sees" the random variable w, but doesn't see the pertinent p at the time that he makes his decision to work. Suppose that the log wage w and log price p obey.

$$w = z + u,$$

$$p = z + v,$$

And Ezu=Ezv=Euv=Eu=Ez=Ev=0. Here z represents neutral movements in the aggregate price level that leave the real wage unaltered. The variates u and v represent factors calling for real wage changes. The worker's linear least squares estimate of (w-p) based on observing w and knowing the first and second moments of all random variables is

$$P[(w-p)|1, w] = a_0 + a_1 w.$$

We have

$$w - p = u - v$$

$$w = z + u.$$

So the normal equations imply

$$a_0 = 0$$

$$a_1 = \frac{E[(u-v)(z+u)]}{E[(z+u)^2]}$$

$$a_1 = \frac{Eu^2}{Ez^2 + Eu^2} \quad .$$

Notice that $0 < a_1 < 1$, and that the greater is Eu^2/Ez^2, the closer to unity is a_1. That makes sense, since the greater is Eu^2/Ez^2, the larger is the fraction of variance in w that is due to variation in the real-wage determining factor u.

The Term Structure of Interest Rates

David Meiselman's[3/] error-learning model of the term structure of interest rates can be described quite compactly and motivated elegantly by using our results on recursive regression. The term structure of interest rates refers to interest rates on assets of similar quality but varying terms to maturity viewed as a function of the yield to maturity. The "yield curve" is a graph of yields to maturity against the maturity. The "yield to maturity," on a bond is defined as the (single) yield that makes the present value of the bond's (expected) stream of payments just equal to the present market price of the bond. The yield to maturity is seen to be equivalent with Keynes's "internal rate of return."

Let R_{nt} be the yield to maturity at time t on a bond that will mature at time t+n. Irving Fisher and John Hicks suggested viewing the n-period yield as an average of the current one-period yield and a sequence of one-period forward rates:

$$(17) \qquad R_{nt} = \frac{1}{n} [R_{1t} + {}_{t+1}F_{1t} + {}_{t+2}F_{1t} + \ldots + {}_{t+n-1}F_{1t}] \qquad n=1, 2, \ldots$$

[3/] David Meiselman, The Term Structure of Interest Rates, Englewood Cliffs, Prentie Hall, 1963. The present description of Meiselman's model is along the lines developed after Meiselman wrote by Mincer, Pye, Diller, Shiller, and Nelson.

where ${}_{t+j}F_{1t}$ is the one-period forward rate that at time t pertains to one-period loans to be made at time t+j and mature at time t+j+1. Equation (17) for n=1, 2, ... actually <u>defines</u> the forward rates ${}_{t+j}F_{1t}$ as functions of the observable yields R_{1t}, R_{2t}, R_{3t}, ... Thus, using (17) we have

$$2R_{2t} = R_{1t} + {}_{t+1}F_{1t}$$

or

$${}_{t+1}F_{1t} = 2R_{2t} - R_{1t}$$

Similarly, we could calculate

$${}_{t+j}F_{1t} = (j+1)R_{j+1,t} - jR_{jt}, \qquad j=1,\ 2,\ \ldots.$$

Now markets in forward loans (i.e., contracts executed at time t for loans to extend between some times t+j to t+k, k>j>0) do not literally exist, as futures markets do in some commodities. Fisher's and Hicks's point was that it was fruitful to decompose a given long rate into the implicit one-period forward rates composing it. Thus, a loan for two periods made at time t is viewed as a one-period (spot) loan made at time t plus a forward commitment entered into at time t to extend the loan for one additional period at time t+1.

So far, all of this has been tautological, since (17) is only a <u>definition</u> of forward rates. Hicks added content to (17) by adopting the expectations hypothesis, asserting that speculators would force forward yields into equality with the spot one-period rates that they expect to hold on the dates to which the forward rates pertain:

$$(18) \qquad {}_{t+j}F_{1t} = \hat{R}_{1,t+j}$$

where $\hat{R}_{1t+j}$ is speculators' forecast of the one-period rate which, as of time t, they expect to prevail at time t+j. Hicks's argument was that unless (18) held, speculators could always increase their expected returns by the appropriate combination of issuing and purchasing debts of various maturities. Thus, suppose that we have the following situation:

$$\left.\begin{aligned} R_{1t} &= .05 \\ R_{2t} &= .04 \end{aligned}\right\} \Rightarrow {}_{t+1}F_{1t} = .03$$

$$\hat{R}_{1t+1} = .05$$

Speculators expect one-period rates to remain stable between period one and two, but the two-period rate is below the one-period rate, indicating that the one-period forward rate is below the one-period spot rate. In this situation, speculators could increase their expected returns, say, by borrowing for two periods (at .04 per period for two periods) and putting the proceeds into a one-period bond the first period (at .05) and another one-period bond the second period (at a return that they expect to be .05). By excuting this transaction, the speculator expects to get a net return of .01 of the amount borrowed each period. (Notice that the speculator need put up no money of his own. Notice also, however, that the speculator is undertaking a risky investment, since the return is not a sure thing--the investor is not in an arbitrage situation where the returns are sure). Hicks entertained the hypothesis that speculators would dominate the market and force (18) to hold.

To make (18) operational, suppose we adopt a version of Muth's hypothesis of "rational expectations" and assume that $\hat{R}_{1t+j}$ is formed as the projection of R_{1t+j} against current and lagged spot rates R_{1t}, R_{1t-1}, ... Then (18) becomes

(19) $$_{t+j}F_{1t} = P[R_{1t+j} | 1, R_{1t}, R_{1t-1}, \ldots].$$

Applying our recursive regression formula (16) we have

(20) $$P[R_{1t+j} | 1, R_{1t+1}, R_{1t}, \ldots] = P[R_{1t+j} | 1, R_{1t}, R_{1t-1}, \ldots] + b_j(R_{1t+1} - P[R_{1t+1} | 1, R_{1t}, R_{1t-1}, \ldots]$$

Substituting (19) into (20) gives

(21) $$_{t+j}F_{1t+1} - {}_{t+j}F_{1t} = b_j(R_{1t+1} - {}_{t+1}F_{1t}).$$

Equation (21) is exactly the "error-learning" model proposed and implemented by Meiselman. Notice that (21) has no random disturbance, so that strictly speaking (21) should fit perfectly--the $\bar{R}^2$ statistic should be unity. In this sense, high values of the $\bar{R}^2$ statistic confirm the theory. Meiselman estimated (21) for annual U.S. data over the period 1901-1954 found positive and statistically signficant b_j's, and found zero constant terms which (21) predicts (though as John Wood and Reuben Kessel pointed out, models other than Meiselman's might also be consistent with the zero intercept). Meiselman found moderately high values for the $\bar{R}^2$ statistics, though they weren't all that close to unity (they ranged from .91 for j=1 to .34 for j=8).

It is straightforward to convert (21) into a regression equation with a disturbance. To illustrate how, suppose that speculators form their expectations at time t by projecting R_{1t+j} on a set $R_{1t}, R_{1t-1}, \ldots, y_t, y_{t-1}, \ldots$, where y_t is some random variable, distinct form R_{1t}, that is useful for forecasting R_{1t+j}. We can thus write

$$_{t+j}F_{1t} = P[R_{1t+j}|\Omega_t]$$

where $\Omega_t = R_{1t}, R_{1t-1}, \ldots, y_t, y_{t-1}, \ldots$. With little additional work, one can deduce the following bivariate version of our recursive learning formula (16):

$$(22)\qquad P[R_{1t+j}|\Omega_{t+1}] - P[R_{1t+j}|\Omega_t] = \gamma_j(R_{1t+1}-P[R_{1t+1}|\Omega_t]) + \beta_j(y_{t+1}-P[y_{t+1}|\Omega_t]),$$

γ_j and β_j being the regression coefficients in the bivariate regression of $R_{1t+j} - P[R_{1t+j}|\Omega_t]$ regressed against $R_{1t+1} - P[R_{1t+1}|\Omega_t]$ and $y_{t+1} - P[y_{t+1}|\Omega_t]$.

Equation (22) is obviously in the same spirit as (16) and says that forecasts are revised according to the "surprising" information in the new observations y_{t+1} and R_{1t+1}. Equation (22) can be written

$$(23)\qquad _{t+j}F_{1t+1} - {}_{t+j}F_{1t} = \gamma_j(R_{1t+1}-P[R_{1t+1}|\Omega_t]) + \beta_j(y_{t+1}-P[y_{t+1}|\Omega_t])$$

Now simply project the right-hand side of the above equation against $R_{1t+1} - P[R_{1t+1}|\Omega_t]$ to get the representation

$$(24)\qquad \gamma_j(R_{1t+1}-P[R_{1t+1}|\Omega_t]) + \beta_j(y_{t+1}-P[y_{t+1}|\Omega_t]) = \phi_j(R_{1t+1}-P[R_{1t+1}|\Omega_t]) + \varepsilon_{t+1}$$

where

$$\phi_j = \frac{\gamma_j \beta_j E(R_{1t+1} - P[R_{1t+1}|\Omega_t])(y_{t+1} - P[y_{t+1}|\Omega_t])}{\beta_j E(R_{1t+1} - P[R_{1t+1}|\Omega_t])^2} ,$$

and where by the orthogonality principle ε_{t+1} is orthogonal to $R_{1t+1} - P[R_{1t+1}|\Omega_t]$. Substituting equation (23) into equation (24) gives

(25) $$_{t+j}F_{1t+1} - {}_{t+j}F_{1t} = \phi_j(R_{1t+1} - {}_{t+1}F_{1t}) + \varepsilon_{t+1} ,$$

which is a regression equation that is in the form of Meiselman's error-learning model. The presence of the random term ε_{t+1} means that there is no implication that equation (25) will bear a high $\bar{R}^2$ statistic.[4/] High values of the $\bar{R}^2$ statistic would indicate that a large proportion of the information useful for forecasting interest rates is included in current and lagged one-period rates.

[4/] Further, notice that we have no restrictions on the signs of ϕ_j; they may be either positive or negative, depending on the various covariances among "surprises" that go into composing ϕ_j.

Notes on the Consumption Function

The literature on the consumption function is primarily addressed to explaining three empirical findings that emerged from early attempts to fit to actual data the simple linear Keynesian consumption function.

(1) $$C = a + bY.$$

For cross-sections where the data on C and Y correspond to n observations on the consumption and income of n households over some short period of time, estimates of (1) typically are characterized by $a > 0$, so that the average propensity to consume (APC) exceeds the marginal propensity to consume (MPC). Similarly, for aggregate time series regressions, where the data on C and Y are economy-wide total consumption and income over a year, estimates of (1) reveal $a > 0$ and an APC > MPC. For example, for annual data for the U.S. for the period 1929-1941, where C is consumption expenditures and Y is disposable income, Ackley[1/] (p. 225) reports the estimated consumption function

$$C_t = 26.5 + .75Y_t.$$

As against the above findings, however, data assembled by Kuznets that extended over the period 1869-1938 and that consisted of (overlapping) ten-year averages of data on aggregate consumption and aggregate disposable income, gave estimates of (1) with b of about .9 and a of about zero. These data were interpreted as indicating that in the very long run, APC = MPC, and consumption is proportional to income.

[1/] Gardner, Ackley, Macroeconomic Theory, MacMillan, 1960.

The tasks of the literature on the consumption function have mainly been:

a) to reconcile the disparity between the time series regressions fitted over short periods, which have APC > MPC, with the proportional (APC = MPC) consumption schedules estimated using Kuznets' data over very long periods of time; and

b) to reconcile the difference between the cross-section regressions that portray APC > MPC with the implications of Kuznets' data.

These notes describe aspects of Milton Friedman's celebrated explanation of these empirical paradoxes.[2/] The treatment here is compatible with Friedman's work, but at some points deviates from being a simple reproduction of it.

The foundation of Friedman's theory is the hypothesis that essentially consumption is proportional to income, measured appropriately. Whether or not the proportionality of consumption and income in Kuznets' data is evidence for that hypothesis is something we shall discuss presently. (Actually, though, Kuznets' data have often been interpreted as lending support to the hypothesis that the true long-run relationship between consumption and income is a proportional one.)

Friedman began with Irving Fisher's theory about consumers' saving. Following Fisher, he posited that the representative household seeks to maximize utility U, where

---

[2/] Milton Friedman, <u>A Theory of the Consumption Function</u>, NBER, 1956.

$$U = U(C_0, C_1, \ldots, C_n)$$

and U(   ) satisfies $U_i > 0$, and is strictly concave; C_i is the household's consumption in period i. The household is assumed to be able to borrow or lend all it desires for i periods at the i-period market determined interest rate R_i. The househould is then supposed to maximize U(   ) subject to the constraint

$$C_0 + \sum_{i=1}^{n} \frac{C_i}{(1+R_i)^i} = Y_0 + \sum_{i=1}^{n} \frac{Y_i}{(1+R_i)^i}$$

where Y_i is the household's income in period i; the constraint thus states that the present value of the household's consumption program must equal the present value of its income stream, i.e., its wealth.

From the assumption that utility is homothetic in consumption at different points in time, Friedman deduced that current consumption is proportional to wealth, the factor of proportionality k depending on the interest rate, among other things.

(2) $C = k(\quad)W$

where $W = Y_0 + \sum_{i=1}^{n} \frac{Y_i}{(1+R_i)^i}$. For several good reasons, Friedman chose to develop the model by at this point introducing the concept of permanent income, which can be defined as the average rate of income that the consumer expects to receive over the rest of his life. Like wealth or present value, permanent income is thus a concept that collapses a stream over time of prospective income into a single summary measure. Permanent income then takes the place of wealth in (2), which is modified to become

(3) $C = \beta(\quad)Y_p$

To make (3) tractable for the purposes of empirical implementation, the dependence of β on the rate of interest and its other determinants is ignored, at least for analyzing the questions to be discussed here, though for other questions the dependence of β on various variables played an important part in Friedman's analysis.

The Cross Section Data

For cross sections, Friedman proposed the model

(4) $C_i = \beta Y_{pi} + u_i$

(5) $Y_i = Y_{pi} + Y_{Ti}$

Here C_i is measured consumption of the ith household, Y_i is measured income of the ith household, while Y_{Ti} is transitory income of the ith household; u_i is the nonsystematic or transitory part of the ith household's consumption. Friedman assumed that Y_{Ti} and u_i both possess zero means

(6) $EY_{Ti} = Eu_i = 0.$

He further assumed the following orthogonality conditions:

(7) $E(Y_{Ti} \cdot u_i) = 0$

(8) $E(Y_{Ti} \cdot Y_{pi}) = 0$

(9) $E(u_i \cdot Y_{pi}) = 0.$

Condition (9) says that equation (4) is a regression equation: that is, the disturbance in (4), which is u_i, obeys the orthogonality condition $Eu_i \cdot Y_{pi} = 0$ which earlier we showed to characterize uniquely the least squares linear regession of C_i on Y_{pi}. Condition (7) states that transitory income and transitory consumption are uncorrelated, that is, are on average unrelated (linearly). Condition (8) is the assumption that permanent income and transitory income are uncorrelated: transitory income is assumed to be randomly distributed with respect to permanent income, in the sense that on average, poor people are as likely to have high (low) transitory income as are rich people.

If data on Y_{pi} were available, (4) could be estimated well by least squares regression, by virtue of the orthogonality condition (9). Indeed, the population linear regression coefficient of C_i against Y_{pi} and 1 is given by

$$b = \frac{E\{(C_i - EC_i)(Y_{pi} - EY_{pi})\}}{E[(Y_{pi} - EY_{pi})^2]}$$

$$= \frac{E\{(\beta(Y_{pi} - EY_{pi}) + u_i)(Y_{pi} - EY_{pi})\}}{E[(Y_{pi} - EY_{pi})^2]}$$

$$= \frac{\beta \text{ var } Y_{pi} + Eu_i \cdot (Y_{pi} - EY_{pi})}{\text{var } Y_{pi}}$$

$$= \beta$$

by virtue of the orthogonality condition (9) and condition (6) on the mean of u_i. Thus, β is the population regression coefficient of C_i against Y_{pi} and 1.

However, data on Y_{pi} are typically not available. What are the consequences of regressing C_i on measured Y_i, as the cross-section studies did, assuming that the model (4)-(9) is correct? The population regression coefficient of C_i against Y_i and 1 is given by

$$h = \frac{E[(C_i-EC_i)(Y_i-EY_i)]}{E[(Y_i-EY_i)^2]}$$

$$= \frac{\text{cov}(C_i,\ Y_i)}{\text{var } Y_i} \quad .$$

We have

$$\text{var } Y_i = \text{var } Y_{pi} + \text{var } Y_{Ti},$$

since the variance of the sum of two uncorrelated random variables equals the sum of their variances. We also have, using (4) and (5),

$$\text{cov}(C_i,\ Y_i) = E\{(C_i-EC)(Y_i-EY_i)\}$$

$$= E\{\beta(Y_{pi}-EY_p)((Y_{pi}-EY_{pi})+Y_{Ti})\}$$

$$= \beta \text{ var } Y_p,$$

since $E(Y_{pi}-EY_p)\cdot Y_{Ti} = 0$ by virtue of assumptions (6) and (8). It follows that

$$\text{(10)} \qquad h = \beta\frac{\text{var } Y_{pi}}{\text{var } Y_{pi}+\text{ var } Y_{Ti}},$$

so that $h < \beta$ so long as var $Y_T > 0$. According to (10), the population value of the linear regression coefficient h is biased downward when taken as an estimate of β, the marginal propensity to consume out of permanent income. We can determine the constant in the population regression of C_i against Y_i as follows. The linear regression line always goes through the means of the variables, so that we have

$$EC_i = k + hEY_i$$

or

$$k = EC_i - hEY_i$$

$$= \beta EY_{pi} - hEY_{pi}$$

$$k = EY_{pi}(\beta - h). \quad (11)$$

Since $\beta - h > 0$ and $EY_{pi} > 0$, we have $k > 0$. Thus, Friedman's model predicts that the cross-section population regressions will yield a positive intercept and an estimated marginal propensity to consume that is less than the marginal propensity to consume out of permanent income.

One way to think of what is going on here is as follows. Let $P[x|1, z]$ be the linear operation "projecting" the random variable x against a constant and the random variable z; so $P[x|1, z]$ just denotes the linear population regression of x against z. For example, using (4) and the orthogonality condition (9), we have

$$P[C_i|1, Y_{pi}] = \beta Y_{pi},$$

which follows because $P[u_i|1,Y_{pi}]=0$. The projection operator is linear in the sense that

$$P[x_i+S_i|1, z_i] = P[x_i|1, z_i] + P[S_i|1, z_i].$$

and

$$P[\alpha x_i | 1, z_i] = \alpha P[x_i | 1, z_i],$$

where α is a scalar. Using these linearity properties of the regression or projection operator, we find from (4) that the regression of C_i on Y_i must obey

(12) $$P[C_i | 1, Y_i] = \beta P[Y_{pi} | 1, Y_i] + P[u_i | 1, Y_i].$$

Now $P[u_i | 1, Y_i] = 0$, by virtue of the orthogonality conditions (7) and (9). This is shown by noting that

$$P[u_i | 1, Y_i] = b_0 + b_1 Y_i$$

where

$$b_1 = \frac{E\{(Y_i - EY_i)(u_i)\}}{E(\{Y_i - Y_i\}^2)}$$

$$b_0 = Eu_i - b_1 EY_i.$$

But we have

$$E\{(Y_i - EY_i)(u_i)\}$$

$$= E\{[(Y_{pi} - EY_{pi}) + Y_{Ti}]u_i\}$$

$$= E\{(Y_{pi} - EY_{pi})u_i\} + EY_{Ti} \cdot u_i$$

$$= 0.$$

Therefore, $b_1=0$, and $b_0=0$, so that

(13) $P[u_i|1, Y_i] = 0.$

To complete (12), we have to calculate

$$P[Y_{pi}|1, Y_i] = \alpha_0 + \alpha_1 Y_i$$

where

$$\alpha_1 = \frac{E\{(Y_i - EY_i)(Y_{pi} - EY_{pi})\}}{E\{(Y_i - EY_i)^2\}}$$

$$\alpha_0 = EY_{pi} - \alpha_1 EY_i.$$

We have

$$\alpha_1 = \frac{E\{((Y_{pi} - EY_{pi}) + Y_{Ti})(Y_{pi} - EY_{pi})\}}{E\{((Y_{pi} - EY_{pi}) + Y_{Ti})^2\}}$$

$$\alpha_1 = \frac{\text{var } Y_{pi}}{\text{var } Y_{pi} + \text{var } Y_{Ti}}$$

$$\alpha_0 = EY_{pi} - \frac{\text{var } Y_{pi}}{\text{var } Y_{pi} + \text{var } Y_{Ti}} EY_i.$$

$$= EY_{pi}\left(1 - \frac{\text{var } Y_{pi}}{\text{var } Y_{pi} + \text{var } Y_{Ti}}\right) = EY_{pi}\left(\frac{\text{var } Y_{Ti}}{\text{var } Y_i}\right)$$

So we have

(14) $$P[Y_{pi}|1, Y_i] = \left(EY_{pi} - \frac{\text{var } Y_{pi}}{\text{var } Y_{Ti} + \text{var } Y_{pi}} EY_i\right).$$

$$+ \frac{\text{var } Y_{pi}}{\text{var } Y_{pi} + \text{var } Y_{Ti}} Y_i.$$

Figure 1

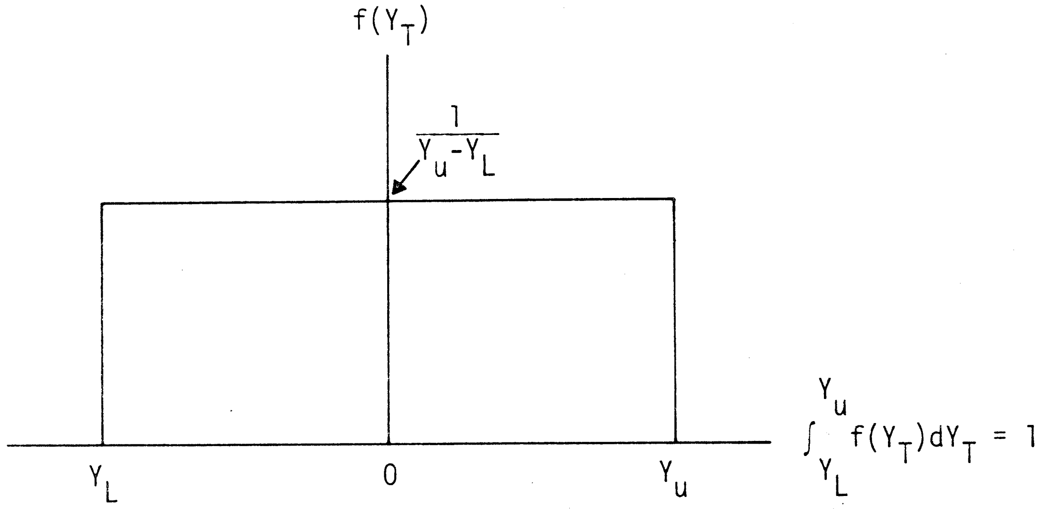


Figure 2

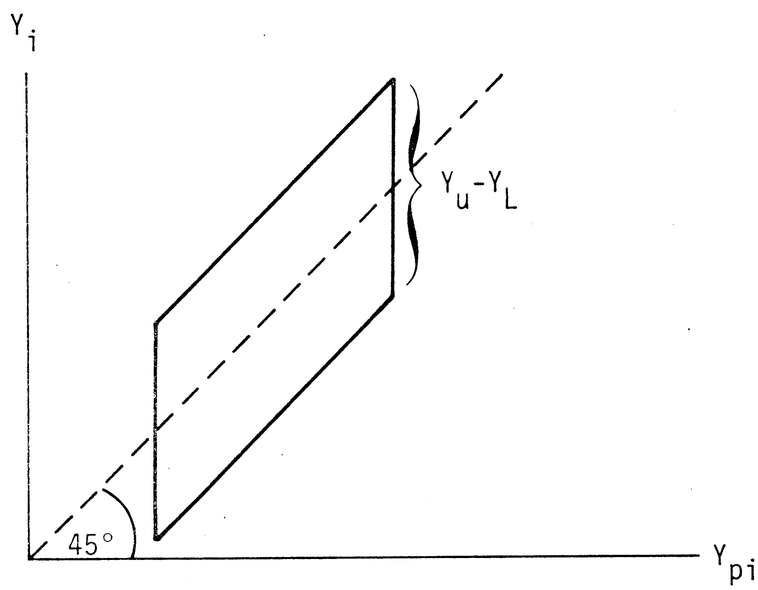


Figure 3

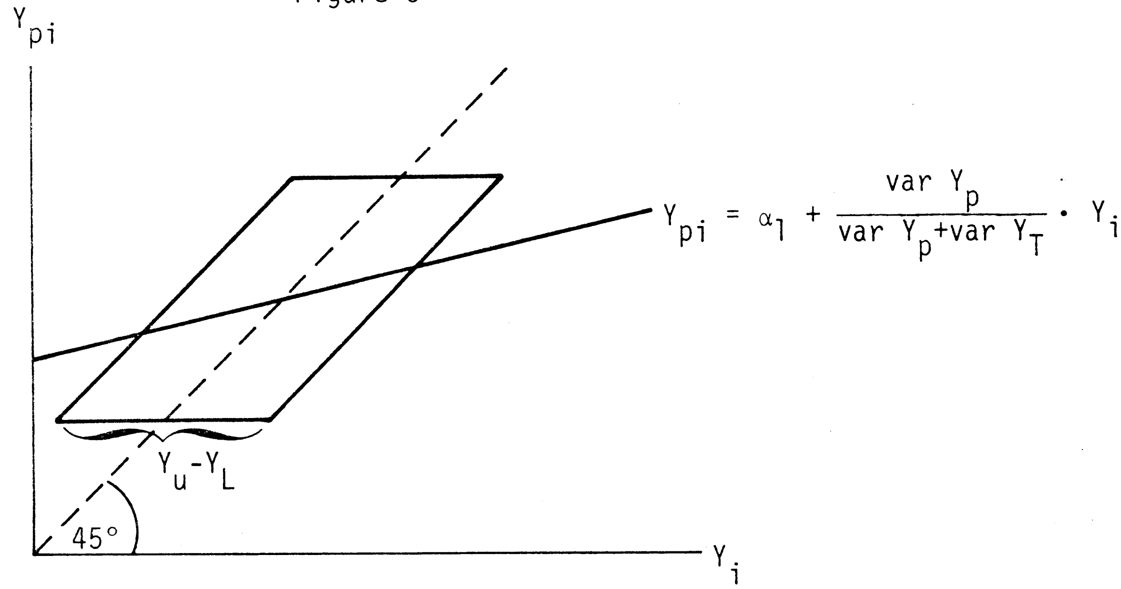


Figure 4

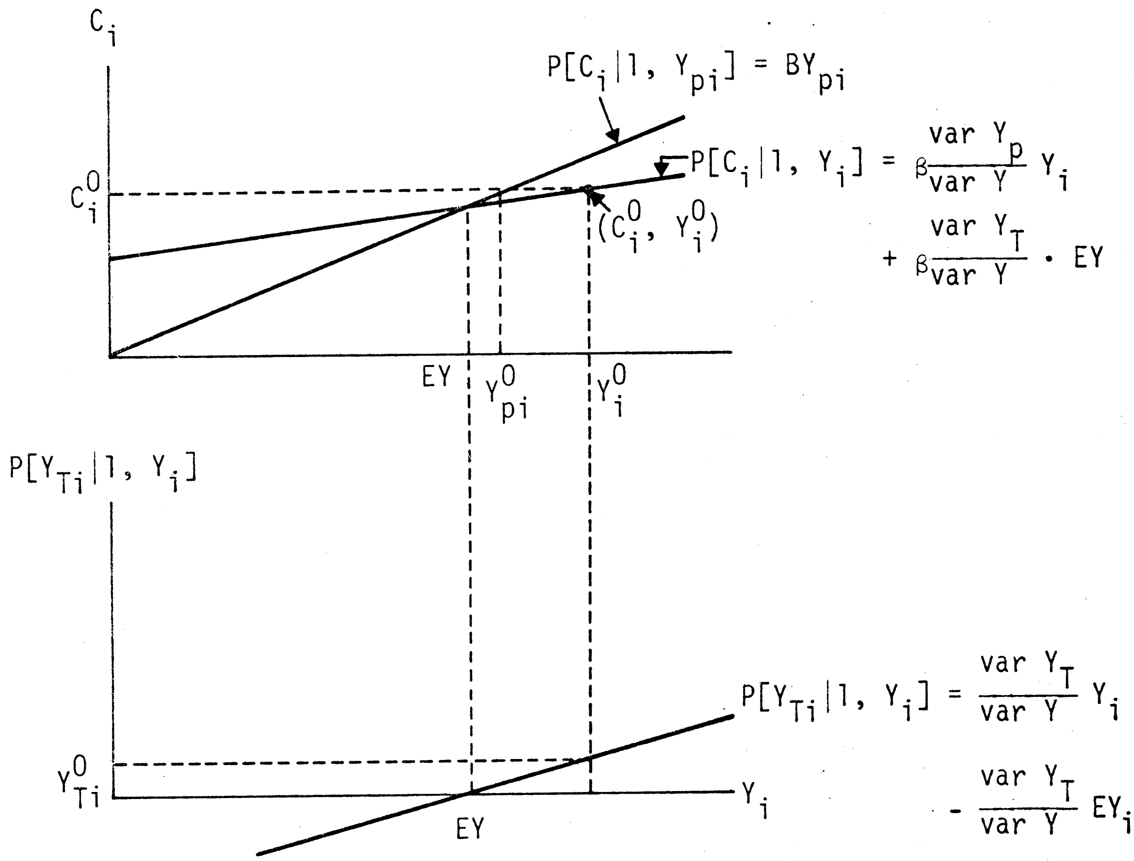
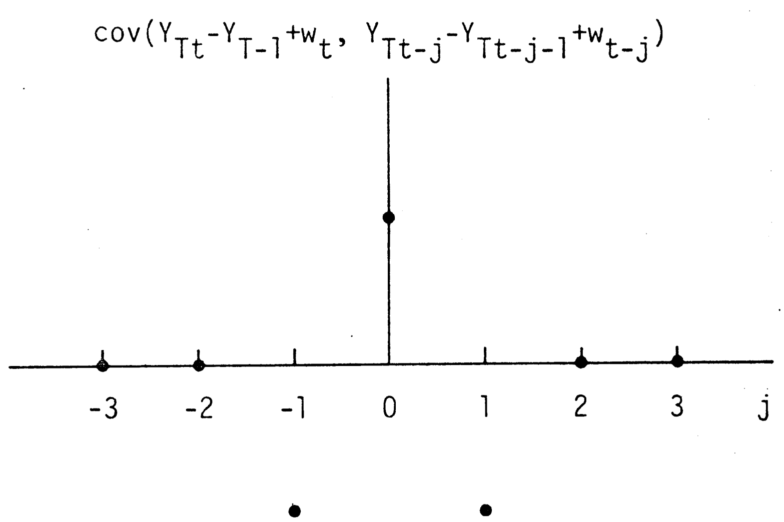


Figure 5



Substituting (13) and (14) into (12) produces our earlier formulas (10) and (11). As reference to (12) shows, the regression of C_i on Y_i has a slope less than β because the regression of the appropriate income concept Y_{pi} has a regression coefficient on the error ridden concept Y_i that is less than unity (so long as var $Y_T > 0$).

Suppose, for example, that Y_{Ti} is distributed according to the uniform distribution

$$\text{Prob}\,[Y_{Ti} < X] = \frac{X - Y_L}{Y_u - Y_L} \qquad \text{for } Y_L \leq X \leq Y_u$$

$$= 1 \qquad X > Y_u$$

$$= 0 \qquad X < Y_L$$

$$Y_L = -Y_u,\; Y_u > 0.$$

The uniform density assumed for Y_T is shown in Figure 1. In Figure 2, the distribution of Y_i and Y_{pi} is indicated. Notice that equation (5) of Friedman's model,

$$(5) \qquad Y_i = Y_{pi} + Y_{Ti}$$

is a regression equation, because the "disturbance" Y_{Ti} is orthogonal to the "regressor" Y_{pi}, implying that the population least squares regression coefficient of Y_i on Y_{pi} equals the value of unity which it bears in (5). However, if we rewrite (5) as

$$Y_{pi} = Y_i - Y_{Ti},$$

we no longer have a regression equation, because the "disturbance" $-Y_{Ti}$ is correlated with the "regressor" Y_i. The least squares regression coefficient of Y_{pi} against Y_i will not equal unity, as we have seen, and as Figure 3 indicates intuitively. Indeed, it is apparent visually from Figure 3 that as var Y_T/var $Y_p \rightarrow \infty$, the slope of the regression line of Y_{pi} against Y_i will go to zero, as our formula (13) indicates. This is so because very little of the variation in Y_i then reflects variation in Y_{pi}.

Now let us put everything together. A consumer with measured income of Y_i will on average have permanent income of

$$(15) \qquad P[Y_{pi}|1, Y_i] = EY_{pi} \cdot \left(\frac{\text{var } Y_{Ti}}{\text{var } Y_i}\right) + \frac{\text{var } Y_p}{\text{var } Y_p + Y_T} \cdot Y_i .$$

On average his transitory income must then be[3/]

$$(16) \qquad P[Y_{Ti}|1, Y_i] = -EY_{pi} \left(\frac{\text{var } Y_{Ti}}{\text{var } Y_i}\right) + \frac{\text{var } Y_T}{\text{var } Y_p + \text{var } Y_T} Y_i .$$

Since the consumer will on average consume according to his permanent income, he will on average consume at the rate

$$\beta P[Y_{pi}|1, Y_i] = \beta[Y_i - P[Y_{Ti}|1, Y_i]].$$

The situation is illustrated in Figure 4. The bottom panel depicts $P[Y_{Ti}|1, Y_i]$ as a linear function going through the point (EY_i, EY_{Ti}), since both means lie on the regression line. The slope of

---

[3/] We derived (16) from (15) by using the facts

$$Y_{Ti} + Y_{pi} = Y_i$$

and

$$P[Y_{Ti}|1, Y_i] + P[Y_{pi}|1, Y_i] = Y_i .$$

the regression line is positive, since on average, transitory income increases when measured income does. Thus, consumers with measured income $Y_i^0 > E(Y_i)$ on average have transitory income $Y_{Ti}^0 > 0$, so that they consume $C_i^0 = \beta(Y_i^0 - Y_{Ti}^0) = \beta Y_{pi}^0$. Consequently, the observation (C_i^0, Y_i^0) lies below the "true" consumption function that relates consumption to permanent income. On the other hand, if $Y_i < EY_i$, then on average transitory income is negative, meaning that measured income on average understates permanent income. The result is that for observations with $EY_i > Y_i$, observations on consumption and measured income will on average lie above the consumption function connecting the permanent magnitudes. The result, then, is to flatten out the consumption function relating consumption to measured income. In this way, Friedman's model reconciles the estimated cross section consumption functions with the hypothesis that consumption is proportional to permanent income.

The Time Series

For the time series, Friedman posited the model

$$C_t = \beta Y_{pt} + U_t \tag{1}$$

where Y_{pt} is aggregate permanent income at time, C_t is aggregate consumption β is the marginal propensity to consume out of permanent income, and U_t is a random disturbance term with mean zero and finite variance. The disturbance term is assumed to obey

$$EU_tY_{ps} = 0$$

for all integer t and s, which implies that (1) is a regression equation.

Since Y_{pt} is not directly observable to make (1) operational requires a model linking permanent income to observable data. To begin, we will assume

(2) $$Y_{pt} = \frac{1}{n} [\hat{Y}_{t+1} + \hat{Y}_{t+2} + \ldots + \hat{Y}_{t+n}]$$

where $\hat{Y}_{t+j}$ is the public's expectation of income at time t+j, based on information available at time t. To complete the model we need a theory about how the public forms the forecasts $\hat{Y}_{t+j}$. We accomplish this by adopting the hypothesis of rational expectations together with a particular statistical model for the income process which John F. Muth[4/] has showed to be compatible with Friedman's assumptions about expectations formation. Income is assumed to be described by the model:

(3) $$Y_{1t} = Y_{1t-1} + w_t + a$$

(4) $$Y_t = Y_{1t} + Y_{Tt}$$

(5) $$E\, w_t = 0 \qquad \text{for all } t$$

(6) $$E\, Y_{Tt} = 0 \qquad \text{for all } t$$

(7) $$E\, w_t\, w_s = \begin{cases} \sigma_w^2 & t=s \\ 0 & t \neq s \end{cases}$$

[4/] J. F. Muth, "Optimal Properties of Exponentially Weighted Forecasts," Journal of the American Statistical Association, 55, (June 1960).

$$(8) \qquad EY_{Tt}\, Y_{Ts} = \begin{cases} \sigma_T^{\,2} & t=s \\ 0 & t\neq s \end{cases}$$

$$(9) \qquad EY_{Tt} \cdot w_s = 0 \quad \text{for all } t,\ s.$$

Here Y_t is measured income, while Y_{Tt} is transitory income. The process w_t is a stationary, serially uncorrelated (equation (7)) random process with mean zero and finite variance; a is a constant representing the trend rate of growth of income; Y_{Tt} is a stationary, serially uncorrelated (condition (8)) random process with mean zero and finite variance. The processes w and Y_T are orthogonal at all lags (condition (9)). According to (3), type 1 or "persisting" income Y_{1t} follows a "random walk" with trend or "drift" a. Measured income equals the persisting type 1 income plus the "white noise" Y_{Tt}.

With the statistical model (3)-(9) in hand, we are now in a position to give content to (2). According to (2) what belongs in the consumption function are households' forecasts of subsequent levels of income. How are we to assume that those forecasts are formed? According to Muth's hypothesis of "rational expectations," those forecasts are posited to be the optimal forecasts of economic and statistical theory. Muth's hypothesis is seen to be an application of the hypothesis of optimizing behavior. To implement the hypothesis requires specifying a statistical model, a set of information assumed to be possessed by the public, and a forecasting criterion to be optimized. Equations (3)-(9) comprise our statistical model. We will assume that as information people use current and lagged values of measured income to forecast future income. We will assume that people form least squares forecasts, that is, forecasts that minimize the mean squared of the forecast error. In sum, we assume that people forecast their future income at time t+j as the projection of Y_{t+j} against current and lagged Y's:

$$P[Y_{t+j}|1, Y_t, Y_{t-1}, \ldots]$$

Using (3) we can write

$$Y_{1t+1} = Y_{1t} + a + w_{t+1}$$

$$Y_{1t+2} = Y_{1t} + 2a + w_{t+1} + w_{t+2}$$

$$\vdots$$

$$Y_{1t+j} = Y_{1t} + ja + w_{t+1} + w_{t+2} + \ldots + w_{t+j}.$$

Since

$$Y_t = Y_{1t} + Y_{Tt},$$

we can write

(10) $$Y_{t+1} = Y_{1t} + a + w_{t+1} + Y_{Tt+1}$$

$$\vdots$$

$$Y_{t+j} = Y_{1t} + ja + w_{t+1} + w_{t+2} + \ldots + w_{t+j} + Y_{Tt+j}.$$

Now since w_t and Y_{Tt} are both serially uncorrelated processes, and since (as we shall see) Y_t is simply a linear combination of current and past w's and Y_T's, it follows that

$$P[w_{t+j}|1, Y_t, Y_{t-1}, \ldots] = 0 \qquad j \geq 1.$$

$$P[Y_{Tt+j}|1, Y_t, Y_{t-1}, \ldots] = 0 \qquad j \geq 1.$$

Substituting the above equalities in the projection of each side of equation (10) on $(Y_t, Y_{t-1}, \ldots)$ gives

$$(11) \qquad P[Y_{t+j}|1, Y_t, Y_{t-1}, \ldots] = P[Y_{1t}|1, Y_t, Y_{t-1}, \ldots] + ja$$

for all $j \geq 1$.

According to equation (11), the optimal forecast of Y_{t+j} conditioned on current and past Y_t's is, apart from the trend term ja, the same for all $j \geq 1$. The identity of the forecasts (apart from the constant trend term) over all horizons j conjures up the notion of estimating a "permanent" level of income. It is in this sense that Muth's model provides a deep rationalization of Friedman's notion of permanent income.

We now have to give an explicit formula for the projection $P[Y_{1t}|Y_t, Y_{t-1}, \ldots]$, which will require a little work.[5/] We begin by adding Y_{Tt} to both sides of (3) to obtain

$$Y_{1t} + Y_{Tt} = (Y_{1t-1}+Y_{Tt-1}) + (Y_{Tt}-Y_{Tt-1}) + w_t + a$$

or

$$(12) \qquad Y_t = Y_{t-1} + (Y_{Tt}-Y_{Tt-1}) + w_t + a.$$

[5/] The derivation here is formally correct but for ease of exposition ignores a technical difficulty that arises because of the fact that our $\{Y_i\}$ process is (borderline) nonstationary. In particular, the variance of Y is not finite, making application of least-squares projection theory a touchy matter. However, by using a suitable limiting argument, the formulas in the text can be shown to hold. The interested reader is directed to Peter Whittle, Prediction and Regulation by Linear Least Squares Methods, Van Nostrand, 1963.

The random variable $(Y_{Tt}-Y_{Tt-1}+w_t)$ in (12) has variance

$$E\{(Y_{Tt}-Y_{Tt-1})+w_t\}^2 = 2 \text{ var } Y_T + \text{var } w.$$

The autocovariances of the random variable $(Y_{Tt}-Y_{Tt-1}+w_t)$ are

$$E[Y_{Tt}-Y_{Tt-1}+w_t,\ Y_{Tt-j}-Y_{Tt-j-1}+w_{t-j}]$$

$$= \begin{cases} -\text{var } Y_T & \text{for } j = \pm 1 \\ 0 & |j| \geq 2. \end{cases}$$

The composite random variable $Y_{Tt} - Y_{Tt-1} + w_t$ thus displays negative first-order serial correlation. The second moments of the composite process are completely characterized by its covariogram, which is simply the covariance of the variable with itself lagged j times graphed against j=0, $\pm$ 1, $\pm$ 2, ... We have established that the covariance of the composite error is

$$\text{cov } (Y_{Tt}-Y_{Tt-1}+w_t,\ Y_{Tt-j}-Y_{Tt-j-1}+w_{t-j})$$

$$= \begin{cases} 2 \text{ var } Y_T + \text{var } w, & j = 0 \\ -\text{var } Y_T, & |j| = 1 \\ 0, & |j| \geq 2. \end{cases}$$

The covariogram is depicted in Figure 5.

It is convenient to replace the composite random process $Y_T - Y_{Tt-1} + \varepsilon_t$ by a single random process that equals and therefore has the same first moments and covariogram as the composite process. In particular, by a theorem of Wold we are assured that there exists a random variable ε_t defined by

(13) $$\varepsilon_t - \lambda\varepsilon_{t-1} = Y_{Tt} - Y_{Tt-1} + w_t$$

where ε_t is a stationary, serially uncorrelated random process with mean zero and variance var ε; λ and var ε are parameters to be determined as follows. The variance of $\varepsilon_t - \lambda\varepsilon_{t-1}$ is var $\varepsilon(1+\lambda^2)$; the covariance between $\varepsilon_t - \lambda\varepsilon_{t-1}$ and lagged values obeys

$$E\{(\varepsilon_t-\lambda\varepsilon_{t-1})(\varepsilon_{t-j}-\lambda\varepsilon_{t-j-1})\}$$

$$= \begin{cases} \mathrm{var}\,\varepsilon(1+\lambda^2) & j=0 \\ -\lambda\,\mathrm{var}\,\varepsilon & |j| = 1 \\ 0 & |j| \geq 2 \end{cases}$$

The parameters λ and var ε are determined to insure equality between the covariogram of $\varepsilon_t - \lambda\varepsilon_{t-1}$ and that of the composite process $Y_{Tt} - Y_{Tt-1} + w_t$. We thus require

$$2\ \mathrm{var}\ Y_T + \mathrm{var}\ w = \mathrm{var}\ \varepsilon\ (1+\lambda^2)$$

$$-\ \mathrm{var}\ Y_T = -\lambda\ \mathrm{var}\ \varepsilon$$

These are two equations that can be solved for λ and var ε as functions of var w and var Y_T. The solution for λ turns out to be

$$\lambda = 1 + \frac{1}{2}\left(\frac{\mathrm{var}\ w}{\mathrm{var}\ Y_T}\right) - \sqrt{\frac{\mathrm{var}\ w}{\mathrm{var}\ Y_T}\left(1 + \frac{1}{4}\frac{\mathrm{var}\ w}{\mathrm{var}\ Y_T}\right)}$$

which will obey $0 \leq \lambda \leq 1$.

Substituting (13) into (12) gives

(14) $$Y_t = Y_{t-1} + \varepsilon_t - \lambda\varepsilon_{t-1} + a$$

which, using lag operators, can be written

$$(1-L)Y_t = (1-\lambda L)\varepsilon_t + a \; .$$

Operating on the above equation with $(1-\lambda L)^{-1}$ gives

$$(15) \qquad \frac{(1-L)}{1-\lambda L} Y_t = \varepsilon_t + \frac{a}{1-\lambda} \; .$$

The left side of the above equation can be written

$$\frac{1}{1-\lambda L} Y_t - \frac{L}{1-\lambda L} Y_t = Y_t + \lambda Y_{t-1} + \lambda^2 Y_{t-2} + \ldots - Y_{t-1}$$

$$- \lambda Y_{t-2} - \ldots$$

$$= Y_t - (1-\lambda)Y_{t-1} - (1-\lambda)\lambda Y_{t-2} - \ldots$$

$$= Y_t - \frac{(1-\lambda)}{1-\lambda L} Y_{t-1} \; .$$

Then (15) can be written

$$(16) \qquad Y_t = (\frac{1-\lambda}{1-\lambda L})Y_{t-1} + (\frac{a}{1-\lambda}) + \varepsilon_t .$$

$$Y_t = (1-\lambda)(Y_{t-1}+\lambda Y_{t-2}+\lambda^2 Y_{t-3}+\ldots) + \frac{a}{1-\lambda} + \varepsilon_t \; .$$

Since ε_t is orthogonal to Y_{t-1}, Y_{t-2}, ... and 1, we have

$$P[Y_t | 1, Y_{t-1}, Y_{t-2}, \ldots] = \frac{1-\lambda}{1-\lambda L} Y_{t-1} + \frac{a}{1-\lambda} \; ,$$

or, shifting time subscripts

$$(17) \qquad P[Y_{t+1} | 1, Y_t, Y_{t-1}, \ldots] = \frac{1-\lambda}{1-\lambda L} Y_t + \frac{a}{1-\lambda} \; .$$

When combined with equation (11) we obtain

(18) $$P[Y_{1t}|1, Y_t, Y_{t-1}, \ldots] = \frac{1-\lambda}{1-\lambda L} Y_t + \frac{a}{1-\lambda} - a.$$

Combining (18) and (11) gives us

(19) $$P[Y_{t+j}|1, Y_t, Y_{t-1}, \ldots] = \frac{1-\lambda}{1-\lambda L} Y_t + \frac{a}{1-\lambda} + (j-1)a,$$

which describes how agents are assumed to forecast income over each horizon.

Since we are assuming rational expectations, we have that the $\hat{Y}_{t+j}$ that belong in (2) obey $\hat{Y}_{t+j} = P[Y_{t+j}|1, Y_t, Y_{t-1}, \ldots]$. Then (2) and (19) imply

$$Y_{pt} = \frac{1-\lambda}{1-\lambda L} Y_t + \frac{a}{1-\lambda} + \frac{a}{n}(1+2+\ldots+n-1)$$

or

(20) $$Y_{pt} = \frac{1-\lambda}{1-\lambda L} Y_t + \frac{a}{1-\lambda} + \frac{a(n-1)}{2}$$

Substituting (20) into (1) gives us

(21) $$C_t = \beta[(\frac{a}{1-\lambda}) + \frac{a(n-1)}{2}] + \beta\frac{(1-\lambda)}{1-\lambda L} Y_t + U_t,$$

which is essentially the time series model that Friedman estimated. It should be emphasized that we have a model of the joint C, Y process, which can be written compactly as equations (21) and (14)

(21) $$C_t = \beta[(\frac{a}{1-\lambda}) + \frac{a(n-1)}{2}] + \frac{\beta(1-\lambda)}{1-\lambda L} Y_t + U_t$$

(14) $$(1-L)Y_t = (1-\lambda L)\varepsilon_t + a$$

Given a model like Friedman's (1) and (2) where expectations are posited to be rational, the consumption-on-income regression given by (21) depends on the nature of the stochastic process assumed to be governing Y, as is testified to by the presence of the parameters a and λ of the Y-process (14) in the consumption-income regression (21). According to the theory of rational expectations, the consumption-income regression (21) will change whenever there is a change in the Y process, as for example will occur if λ or a changes in (14). Of course, the change in λ would in turn be attributable to a change in the ratio of the variance of w to the variance of Y_T.

Simulating the Model: The Response to an Unexpected Increase in Income

Suppose that (21) and (14) prevail with fixed a and λ. We can simulate the model in an instructive way by investigating the response of consumption and income to an unexpected change in income (an "innovation" in income). The unexpected part of income is simply ε_t, as can be seen by recalling that

$$Y_t = \frac{1-\lambda}{1-\lambda L} Y_{t-1} + \frac{a}{1-\lambda} + \varepsilon_t$$

and

$$P[Y_t | 1, Y_{t-1}, \ldots] = \frac{1-\lambda}{1-\lambda L} Y_{t-1} + \frac{a}{1-\lambda}$$

Therefore,

$$Y_t - P[Y_t | 1, Y_{t-1}, \ldots] = \varepsilon_t.$$

For simplicity, assume that a=0, and write (8) and (21) as

$$Y_t = \frac{1-\lambda L}{1-L} \varepsilon_t$$

$$C_t = \frac{\beta(1-\lambda)}{1-\lambda L} Y_t + U_t$$

$$= \frac{\beta(1-\lambda)}{1-\lambda L} \frac{(1-\lambda L)}{(1-L)} \varepsilon_t + U_t$$

$$= \frac{\beta(1-\lambda)}{(1-L)} \varepsilon_t + U_t.$$

Writing out these expressions for C_t and Y_t we have

(22) $$Y_t = \varepsilon_t + (1-\lambda)\varepsilon_{t-1} + (1-\lambda)\varepsilon_{t-2} + \ldots$$

(23) $$C_t = \beta(1-\lambda)(\varepsilon_t+\varepsilon_{t-1}+\varepsilon_{t-2}+\ldots) + U_t$$

Equation (22) shows that a random unexpected income of ε_t causes Y_t to increase by ε_t and can be expected to cause Y in each subsequent period to jump by $(1-\lambda)\varepsilon_t$. Thus, an unexpected jump in income of ε_t causes a jump in permanent income of $(1-\lambda)\varepsilon_t$. Equation (23) shows that the jump in ε_t causes C_t to jump by $\beta(1-\lambda)\varepsilon_t$, which equals β times the change in permanent income. Equation (23) indicates that consumption in all subsequent periods can also be expected to increase by $\beta(1-\lambda)\varepsilon_t$, so that the unexpected change in income of ε_t can be expected to set off a permanent change in consumption of $\beta(1-\lambda)\varepsilon_t$.

The preceding discussion indicates how λ measures the fraction of an unexpected change in income that the public imputes to transitory income Y_{Tt}, while $(1-\lambda)$ measures the fraction that the public imputes to "persisting" income Y_{1t}. Our formula for λ is

$$\lambda = 1 + \frac{1}{2}\left(\frac{\text{var } w}{\text{var } Y_T}\right) - \sqrt{\frac{\text{var } w}{\text{var } Y_T}\left(1+\frac{1}{4}\frac{\text{var } w}{\text{var } Y_T}\right)}.$$

By differentiating with respect to var w/var Y_T, it can be shown that λ is a decreasing function of var w/var Y_T. Notice that as (var w/var Y_T)

approaches zero, λ approaches 1, which is consistent with our interpretation of λ as the fraction of unexpected changes in income that rational agents impute to transitory income.

Notice that in "simulating" the model, we have taken care to preserve the stochastic structure formed by the joint model (21)-(14), by in effect drawing random terms from the distribution governing the ε's. We did not simply impose an arbitrary path on income, and then hold (21) fixed in spite of the implicit change in the Y process that this entails. Such simulations, which in effect ignore the dependency of the parameters in (21) on the parameters of the income process (14), are commonly used to study Friedman's model, but are clearly faulty devices for doing so.

Simulating the Model: Response to a Change in the Trend Rate of Growth of Income

Suppose that at some point in time there occurs a widely known once and for all change in the trend rate of growth of income, as represented by a jump in a with λ and the variance of ε fixed. Under this circumstance the intercepts in (21) and (14) will change once and for all, with no other changes.

Simulating the Model: A Random Tax Credit[6/]

Suppose the government institutes a policy of imposing on consumers what they properly perceive as being a random sequence of tax credits, lump-sum taxes, x_t, so that (14) must be modified to be

[6/] This is one of the examples analyzed in the important paper by Robert E. Lucas, Jr., "Econometric Policy Evaluation: A Critique," Journal of Monetary Economics, Supplement, 1976.

(14') $(1-L)Y_t = (1-\lambda L)\varepsilon_t + x_t + a$

where x_t is a stationary, serially uncorrelated random process with mean zero. We assume that x is orthogonal to w and to Y_T. By repeating our earlier analysis, it is straightforward to show that (14') can be rewritten as

$$(1-L)Y_t = (1-\lambda' L)\varepsilon_t' + a$$

where

$$\lambda' = 1 + \frac{1}{2}\left(\frac{\text{var } w}{\text{var } Y_T + \text{var } x}\right) - \sqrt{\frac{\text{var } w}{\text{var } Y_T + \text{var } x}\left(1+\frac{1}{4}\left(\frac{\text{var } w}{\text{var } Y_T + \text{var } x}\right)\right)}.$$

It is clear that $\lambda' > \lambda$, so that the effect of instituting the policy is to increase λ in (21) and (14). This is a natural outcome, since the policy in effect increases the variance of transitory income relative to the variance of innovations in persisting income, w. Notice that the result of instituting the policy is to change the parameters of the statistical model (21) - (14) that links consumption to income. Thus, it would be a mistake to simulate the effects of such a policy by simply plugging the values of a Y_t sequence generated under the new policy into (14) with the parameter λ that prevailed under the <u>old</u> policy, the value of λ that would (in large samples) be recovered by econometric analysis of historical time series over the period of the old policy.

Compatability of Friedman's Model with the Time Series Regressions

Friedman's model is widely regarded as reconciling the discrepancies between the consumption-income regressions on annual data with those using Kuznets' ten-year overlapping decade averages. Indeed,

Table 1--Kuznets' Data

Decade	National Income (billions of 1929 dollars)	Consumption Expenditures (billions of 1929 dollars)
1869-78	9.3	8.1
1874-83	13.6	11.6
1879-85	17.9	15.3
1884-93	21.0	17.7
1889-98	24.2	20.2
1894-1903	29.8	25.4
1899-1908	37.3	32.3
1904-13	45.0	39.1
1909-18	50.6	44.0
1914-23	57.3	50.7
1919-28	69.0	62.0
1924-33	73.3	68.9
1929-38	72.0	71.0

Source: S. Kuznets, National Product Since 1869, (National Bureau of Economic Research, 1946), p. 119.

Table 2
Regressions with Kuznets' Data

Regressions (1) and (2) exclude the observation for 1929-38.

(1) $\bar{C}_t = \underset{(.67)}{-1.72} + \underset{(.016)}{.928}\,\bar{Y}_t$, d.w. = .77, $\bar{R}^2 = .997$

(2) $\bar{C}_t = \underset{(.57)}{-1.28} + \underset{(.082)}{1.137}\,\bar{Y}_t - \underset{(.49)}{1.27}t$, d.w. = 1.65, $\bar{R}^2 = .998$

Regressions (3) and (4) include the observation for 1929-38.

(3) $\bar{C}_t = \underset{(1.09)}{-2.48} + \underset{(.024)}{.958}\,\bar{Y}_t$ d.w. = .53, $\bar{R}^2 = .993$

(4) $\bar{C}_t = \underset{(1.17)}{-2.43} + \underset{(.172)}{.989}\,\bar{Y}_t - \underset{(1.01)}{.18}t$ d.w. = .57, $\bar{R}^2 = .992$

the "proportionality" of the regressions on Kuznets' regression data is often taken as evidence in favor of Friedman's hypothesis that the true relation between consumption and permanent income is one of proportionality. Further, the slope of the regression on Kuznets' data seems to be regarded as a good estimate of the marginal propensity to consume out of permanent income or the "long run" marginal propensity to consumer. Here we investigate whether the model consisting of (14) and (21) does in fact reconcile the various time series consumption functions. Our method for doing this is straightforward: we assume that (14) and (21) prevail, and then calculate the implied simple regressions of consumption on income for both the annual data and decade-averages.

Table 1 records Kuznets' overlapping decade averages of consumption and income. Table 2 reports least squares regressions of average consumption on average income, both excluding and including a trend, and excluding and including the last observation corresponding to the decade of the Great Depression. It is the regressions without the trend terms that are widely regarded as recovering a good estimate of β as the coefficient on $\overline{Y}_t$. We report the regressions including the trend term to highlight how excluding it is required to deliver a coefficient on $\overline{Y}_t$ that seems plausible as an estimate of β. The calculations below provide reasons for expecting that the regression excluding the trend will provide a better estimate of β. However, those calculations also indicate that if Friedman's consumption-income model consisting of equations (14) and (21) is correct, then the regression including the trend term can be expected to underestimate β. The point estimates in Table 2 are not consistent with that prediction.

The calculations in this section are intended to illustrate a rigorous method for evaluating descriptive interpretations such as the following one offered by Daniel Suits:

> In the discussion of long-run and short-run effects two things are sometimes confused: the nature of the problem under investigation, and the nature of the data employed. It is possible to use quarterly data and still analyze a very long-run consumption function. The data set a lower limit to the "length of run" that can be investigated--the Kuznets' estimates for decades cannot be used to investigate quarterly variations in consumption, but they do not, of themselves, set an upper limit. A regression fitted to annual, quarterly, or even monthly data for the period 1865 to the present would yield results essentially no different from that obtained from decade averages. When we use a time span covering nearly a hundred years, the regression analysis is going to be most sensitive to the big overall changes, to the general drift of the data and not to the relatively minor differences between one year and the next.[7/]

We proceed to calculate the simple regression coefficient that would obtain on n-period average data if Friedman's model consisting of equations (14) and (21) were correct. To simplify the calculations, we will work in first differences and write equations (14) and (21) as

$$(22) \qquad y_t = (1-\lambda L)\varepsilon_t + a$$

$$c_t = \frac{\beta(1-\lambda)}{1-\lambda L} y_t + u_t,$$

or

$$(23) \qquad c_t = \beta(1-\lambda)\varepsilon_t + \beta a + u_t$$

where $c_t = C_t - C_{t-1}$, $y_t = Y_t - Y_{t-1}$, $u_t = U_t - U_{t-1}$.

7/ Suits, Daniel B., "The Determinants of Consumer Expenditure: A Review of Present Knowledge," Impacts of Monetary Policy, Commission on Money and Credit, 1963, pp. 34-35.

Now consider forming n-period moving averages of y_t and c_t:

$$\bar{y}_t = \frac{1}{n}(1+L+\ldots+L^{n-1})y_t$$

$$\bar{c}_t = \frac{1}{n}(L+L+\ldots+L^{n-1})c_t.$$

Taking n-period moving averages on both sides of (22) and (23) gives

$$\bar{y}_t = a + \frac{1}{n}(\varepsilon_t+\ldots+\varepsilon_{t-n-1}-\lambda\varepsilon_{t-1}-\ldots-\lambda\varepsilon_{t-n})$$

(22') $$\bar{y}_t = a + \frac{1}{n}(\varepsilon_t+(1-\lambda)\varepsilon_{t-1}+\ldots+(1-\lambda)\varepsilon_{t-n-1}-\lambda\varepsilon_{t-n})$$

(23') $$\bar{c}_t = \beta a + \frac{\beta(1-\lambda)}{n}[\varepsilon_t+\varepsilon_{t-1}+\ldots+\varepsilon_{t-n-1}] + \bar{u}_t .$$

Since successive ε's are orthogonal, we have from (22') that the variance of $\bar{y}_t$ is

$$\sigma_{\bar{y}_t}^{\ 2} = \frac{\sigma_\varepsilon^{\ 2}}{n^2}[1+\lambda^2+(n-1)(1-\lambda)^2]$$

$$= \frac{\sigma_\varepsilon^{\ 2}}{n^2}[1+\lambda^2+(n-1)(1-2\lambda+\lambda^2)]$$

$$\sigma_{\bar{y}_t}^{\ 2} = \frac{\sigma_\varepsilon^{\ 2}}{n^2}[n(1+\lambda^2)-2\lambda(n-1)].$$

The covariance between $\bar{y}_t$ and $\bar{c}_t$ is calculated by using (22') and (23') to calculate $\bar{y}_t - E\bar{y}_t$ and $\bar{c}_t - E\bar{c}_t$, multiplying, and taking expected values:

$$\sigma_{\bar{c}\,\bar{y}} = \sigma_\varepsilon^{\ 2}\,\frac{\beta(1-\lambda)}{n^2}[1+(n-1)(1-\lambda)]$$

$$\sigma_{\bar{c}\,\bar{y}} = \sigma_{\varepsilon}^{2}\,\frac{\beta(1-\lambda)}{n^{2}}\,[n-(n-1)\lambda].$$

The studies using Kuznets' ten-year averaged data in effect calculated the simple regression through the origin of $\bar{c}_t$ on $\bar{y}_t$ for n chosen to be ten (with annual data). That is they presented the regression

(24) $\bar{c}_t = \gamma\bar{y}_t$ + residual

Since we are working here with first differences, computing a regression of the level of the averaged C on the averaged Y and a constant term with no trend term corresponds to running the regression on first differences through the origin.[8/] Before considering regression (24) it is interesting to analyze the regression with an intercept term,

(25) $\bar{c}_t = \delta\bar{y}_t$ + constant + residual.

The counterpart to (25) is a regression of the level of averaged C on the level of averaged Y, a constant, and a trend. The population value of δ is given by

---

[8/] Suppose that for levels we have

$$X_t = a + bt + cZ_t + r_t,$$

r_t a residual orthogonal to Z at all lags. First differencing gives

$$(x_t - x_{t-1}) = b(t-t+1) + c(z_t - z_{t-1}) + (r_t - r_{t-1})$$

or

$$x_t - x_{t-1} = b + c(z_t - z_{t-1}) + (r_t - r_{r-1}).$$

Thus, the slope c of the regression remains unchanged, while the coefficient on the trend in the level regression becomes the constant term in the regression in first differences.

$$\delta = \frac{\sigma_{\overline{cy}}}{\sigma_{\bar{y}}^2} .$$

Using our formulas for $\sigma_{\overline{cy}}$ and $\sigma_{\bar{y}}^2$ we have

$$(26) \qquad \delta = \frac{\beta(1-\lambda)(n-\lambda(n-1))}{[(1+\lambda^2)n-2\lambda(n-1)]} .$$

How closely does δ approximate β? For n sufficiently large, δ approximates β closely, since

$$\lim_{n\to\infty} \frac{(n-\lambda(n-1))}{(1+\lambda^2)n-2\lambda(n-1)} = \lim_{n\to\infty} \frac{(1-\lambda(\frac{n-1}{n}))}{(1+\lambda^2)-2\lambda(\frac{n-1}{n})}$$

$$= \frac{1-\lambda}{1+\lambda^2-2\lambda} = \frac{1}{1-\lambda} .$$

Thus, we have

$$\lim_{n\to\infty}\delta = \beta.$$

For n=10 and various values of λ, Table 3 reports values of the bias factor

$$(1-\lambda) \left(\frac{n-\lambda(n-1)}{(1+\lambda^2 n-2\lambda(n-1)}\right)$$

which is associated with taking δ as an estimate of β. For λ=.3, the value Friedman found for the annual time series, the bias factor is .929, which is substantial.

For regressions on levels that do not include a trend, (24) is the corresponding model in terms of first differenced data. The population value of the slope coefficient γ is given by

$$\gamma = \frac{\sigma_{\overline{cy}} + E(\bar{c})E(\bar{y})}{\sigma_{\bar{y}}^2+(E(\bar{y}))^2} .$$

Table 3

λ	$(1-\lambda)\left(\frac{10-9\lambda}{(1+\lambda^2)\cdot 10-2\lambda\cdot 9}\right)$
0	1.0
.1	.987
.2	.965
.3	.929
.4	.873
.5	.786
.6	.657
.7	.483
.8	.280
.9	.100
1.0	.0

From (22) and (33) we have that $Ey_t=a$, $Ec_t=\beta a$. Since taking moving averages does not change means, we have that

$$E\overline{y}_t = a$$

$$E\overline{c}_t = \beta a.$$

Consequently, we have that the population parameter γ obeys

$$\gamma = \frac{\sigma_{\overline{cy}} + \beta a^2}{\sigma_{\overline{y}}^2 + a^2} = \frac{\frac{\sigma_{\overline{cy}}}{a^2} + \beta}{\frac{\sigma_{\overline{y}}^2}{a^2} + 1}$$

For fixed values of $\sigma_{\overline{cy}}$ and $\overline{y}^2$, we have that

$$\lim_{a^2\to\infty}\gamma = \beta,$$

so that as the trend term a becomes relatively more and more important, γ approaches β. Using our expressions for $\sigma_{\overline{cy}}$ and $\sigma_{\overline{y}}^2$, we can express γ as

$$\gamma = \frac{\frac{\sigma_\varepsilon^2}{n^2}\,[\beta(1-\lambda)[n-\lambda(n-1)]]+\beta a^2}{\frac{\sigma_\varepsilon^2}{n^2}\,[(1+\lambda^2)n-2\lambda(n-1)]+a^2}$$

or

$$(27) \qquad = \frac{\dfrac{\beta(1-\lambda)(1-\lambda(\frac{n-1}{n}))}{(1-\lambda^2-2\lambda(\frac{n-1}{n}))} + \dfrac{n\beta a^2}{\sigma_\varepsilon^2(1+\lambda^2-2\lambda(\frac{n-1}{n}))}}{1+\dfrac{na^2}{\sigma_\varepsilon^2(1+\lambda^2-2\lambda(\frac{n-1}{n}))}} = \frac{\delta + \dfrac{n\beta a^2}{\sigma_\varepsilon^2(1+\lambda^2-2\lambda(\frac{n-1}{n}))}}{1+\dfrac{na^2}{\sigma_\varepsilon^2(1+\lambda^2-2\lambda(\frac{n-1}{n}))}}$$

Holding β, λ, σ_ε^2, and a fixed, we have

$$\lim_{n\to\infty}\gamma = \frac{\dfrac{\beta(1-\lambda)(1-\lambda)}{(1-\lambda)^2} + \dfrac{n\beta a^2}{\sigma_\varepsilon^2(1-\lambda)^2}}{1 + \dfrac{na^2}{\sigma_\varepsilon^2(1-\lambda)^2}}$$

$$= \frac{\beta\left(1+\dfrac{na^2}{\sigma_\varepsilon^2(1-\lambda)^2}\right)}{1+\dfrac{na^2}{\sigma_\varepsilon^2(1-\lambda)^2}}$$

$$= \beta,$$

so that for averages sufficiently long, the slope γ does approximate β well. Expression (27) shows that γ approximates β better the larger is n, the larger is a, and the smaller is σ_ε^2.

Comparing (26) and (27) for n=1 and n=10 permits evaluating the passage by Suits quoted earlier. With n=1, (26) gives

$$\delta = \frac{\beta(1-\lambda)}{(1+\lambda^2)} ,$$

which is the population slope of a regression of the one-period level C_t against a constant and a trend. Notice that for $\lambda > 0$ this value of δ is less than $\beta(1-\lambda)$, which is often interpreted as the one-period marginal propensity to consumer. For n=1, (27) gives,

$$\gamma = \frac{\beta(1-\lambda)\sigma_\varepsilon^2+\beta a^2}{\sigma_\varepsilon^2(1+\lambda^2)+a^2} ,$$

which is the population slope of a regression of the one-period level of C_t against a constant and the level of Y_t. Clearly, for sizable values of λ, δ for n=1 is very much smaller than δ for n=10. Whether γ for n=1 is close to γ for n=10 depends critically on the ratio of the income-innovation variance σ_ε^2 to the income trend parameter a. The smaller is this ratio, the closer will γ for n=1 be both to β and to γ for n=10. The remarks for Suits are thus approximately valid only under suitable restrictions on σ_ε^2 and a. Further, their validity is crucially dependent on excluding a trend term from the regressions in question.

The preceding calculations provide a rigorous framework for evaluating the claim that the regressions on Kuznets' data estimate the marginal propensity to consume out of permanent income. The calculations indicate that for sizable λ's, the presence of a strong trend in income (a large a) <u>and</u> the omission of a trend term in the regressions on Kuznets' data are essential elements in recovering a good estimate of β.

It is still perhaps an open question whether the trend term a in fact is big enough relative to the variance of unexpected income, σ_ε^2, to make γ a good approximation to β for ten-period average data.

It is useful to note that for one-period regressions of the form (25), we have found that $\delta = \beta(1-\lambda)/(1+\lambda^2)$. The constant in this regression, say k, will obey

$$k = Ec_t - \delta E_y$$

$$= \beta a - \delta a$$

$$= a(\beta - \beta(1-\lambda)/(1+\lambda^2))$$

$$k = a\beta(\frac{\lambda(1+\lambda)}{1+\lambda^2} > 0$$

so long as λ, α, β all exceed zero.

Therefore, the model predicts that if $a > 0$, the annual regression of consumption on current income, a constant, and a linear trend will have a positive coefficient on trend. The apparent marginal propensity to consume $\beta(1-\lambda)/(1+\lambda^2)$ will be much lower than β for sizable values of λ, but the consumption function will appear to be drifting upward with the passage of time,

$$C_t = k_0 + kt + \frac{\beta(1-\lambda)}{(1+\lambda^2)} Y_t + \text{residual}_t.$$

Interestingly enough, Smithies used a regression of this form in one of the earliest studies of the time series consumption schedule. He found k to be positive and statistically significant.[9/]

Aspects of reconciling Friedman's model with the time series seem to remain unresolved, particularly whether or not the model (14)-(21) predicts the pattern of estimates made using Kuznets' data. However, to test the model (14) - (21) in a statistically powerful way, it is not really appropriate to proceed in the piecemeal fashion of checking whether the model appears to rationalize regressions that various researchers have calculated on the basis of various kinds of time series data. We conclude our discussion of consumption by indicating briefly how one can go about testing (14) - (21).

[9/] Smithies offered a theory for why b might be positive, though quite a different one from the model (14), (21) which leads us to predict $k > 0$ for $a > 0$.

Testing the Model

There are three aspects of the model of the (C, Y) process formed by (21) and (14) that bear testing. First, there is the adequacy of (14) as a model for the time series income. Second, there is the question of whether the restrictions across (14) and (21) on the parameters seem to hold; that is, (14) and (21) are restricted to share the common parameters a and λ. Third, there is the assumption $EU_tY_s = 0$ for all t, s, the assumption of strict econometric exogeneity of Y in (21).

To illustrate briefly one way to proceed, notice that (21) (14) is a special case of the model

$$
\begin{aligned}
C_t &= k_0 + \sum_{i=-\infty}^{\infty} h_i Y_{t-i} + U_t \\
Y_t &= \alpha + \sum_{i=1}^{\infty} v_i Y_{t-i} + \sum_{i=1}^{\infty} w_i C_{t-i} + \varepsilon_t ,
\end{aligned}
\tag{28}
$$

a special case with certain restrictions on and across the h_i's , v_i's, and w_i's. A reasonable way to proceed is first to estimate (28) by maximum likelihood under the restrictions imposed in the special case of (21) and (14). Then by maximum likelihood estimate (28) under a less restrictive parameterization, i.e., a choice of v_i's, h_i's, and w_i's free enough to include (14), (21) as a special case. An asymptotically valid test of the adequacy of specification (14), (21) is provided by a likelihood ratio statistic than can be computed from the values of the likelihood function attained under these two parameterizations. By doing this, each of the three aspects of the model mentioned above can be tested.

Such tests have not been implemented for Friedman's consumption-income model.

The Phillips Curve

The "Phillips curve" alludes to a negative correlation between wage inflation and unemployment often thought to have been originally spotted by Phillips in the British data, though Irving Fisher and business cycle analysts at the National Bureau of Economic Research and elsewhere had remarked about the correlation long before. From the point of view of the nonrandom classical model, the observed Phillips curve is a paradox, since that model asserts that things that cause inflation, such as growing deficits and high rates of money creation, will leave "real" variables such as unemployment and real GNP unaffected. Any evidence that suggests an influence running from higher aggregate demand to higher real GNP and lower unemployment (rather than to higher prices only) seems to contradict the classical model as we have formulated it. More generally, such evidence seems to contradict any general equilibrium model in which agents' decisions about real economic variables are homogenous of degree zero in nominal magnitudes, as a large body of economic theory predicts.

The initial response of macroeconomists to Phillips' findings was to accept the correlation that Phillips had found as a relation suitable for including in a macroeconometric model--a big jump without first having a well developed theory of that relation. The Phillips curve, expressing the rate of change of wages as a function of unemployment, fit very well into the Keynesian model because it seemed to provide a convenient recursive device for making wages endogenous to the Keynesian model over time, although perhaps still fixed at a point in time (think of the Phillips curve in continuous time as reading $\dot{w}/w = f(U)$, so that

$\dot{w}$ is endogenous at each moment, even while the level of w is fixed). The Phillips curve was widely interpreted as depicting a tradeoff between inflation and unemployment along which policy makers could select a point through suitable monetary and fiscal policy.

The last decade has seen an increasing amount of dissatisfaction with the preceding use of the Phillips curve. Ultimately, the source of that dissatisfaction is the failure of estimated Phillips curves to remain stable over time. The apparent tradeoff between inflation and unemployment has worsened in most western countries in the last decade, as inflation rates have risen. Largely in response to this phenomenon, a body of theoretical work has emerged in an attempt to explain how a Phillips curve could arise and to what extent it represents a tradeoff that policy makers can exploit. Important work in this area has been done by Friedman, Phelps, Alchian, Gordon and Hynes, and Lucas and Prescott.

These pages describe Lucas's model of the Phillips curve.[1/] Lucas's model, and much of the other work in this field, embodies the "natural rate hypothesis," which amounts merely to asserting that agents' decisions depend only on relative prices. Within the confines of such a hypothesis, if one is to explain why high inflation and high nominal aggregate demand seem to induce high aggregate output, it is necessary to construct an operational model of "money illusion." Lucas, in effect, constructed a simple model of "money illusion," one compatible with rational, optimizing behavior. As Lucas put it:

> "All formulations of the natural rate hypothesis postulate rational agents whose decisions depend on relative prices only, placed in an economic setting where they cannot distinguish relative from general price movements."

Lucas supposes that suppliers of a single good are located in a large number of physically separated competitive markets. Demand is distributed unevenly across markets; so that prices of the one good vary across markets; the good is perishable and there is no trading across markets. All markets are identical except for their state of demand. For that reason, we can, if we choose, index markets by a variable measuring their state of demand, z. The variable z is a random variable with characteristics to be described shortly, which we also use to index the markets (if two markets have the same z, they behave identically).

Agents are assumed to know the first and second moments of all probability distributions. This gives them the information needed to form linear least squares projections of random variables they don't know on the random variables they do know. We implement the hypothesis of rational expectations by assuming that agents' expectations about unknown random variables equal the linear least squares projections on certain information sets to be specified.

Supply in market z is assumed to be governed by

$$y_t(z) = \gamma(p_t(z) - P[p_t|I_t(z)]) + \lambda y_{t-1}(z) \qquad (1)$$

$$\gamma > 0 \quad , 0 < \lambda < 1$$

where $y_t(z)$ = logarithm of supply in market z

$p_t(z)$ = logarithm of price in market z (assumed given to the suppliers who are price-takers)

p_t = average economy-wide logarithm of price (the average across markets of the $p_t(z)$'s

$P[p_t|I_t(z)]$ = the projection of p_t on information available in market z at time t

$I_t(z)$ = information available at time t in market z.

According to (1), supply in market z responds directly to the gap between the current price in market z and the forecast $P[p_t|I_t(z)]$ of average economy-wide price made by agents in market z. Agents in z are assumed not to know p_t because at time t they see only the price in their own market $p_t(z)$. For that reason, they have to forecast p_t by projecting it on information they do have, an information set to be specified shortly. Equation (1) depicts agents as responding to what they perceive to be increases in the relative price $p_t(z) - P[p_t|I_t(z)]$, but as failing to respond to what are perceived as general increases in the price level, i.e., those that leave $p_t(z) - P[p_t|I_t(z)]$ unaltered. The term $\lambda y_{t-1}(z)$ is added on to account for the possibility that supply responds also to lagged perceived relative price changes. We shall say more about this shortly.

To complete the model, we have to specify the information set $I_t(z)$. We initially assume that $I_t(z)$ consists of two components. First, $I_t(z)$ of course includes $p_t(z)$, since agents in market z see the price facing them at t. We assume that $p_t(z)$ is the only current information that agents receive. Second, $I_t(z)$ includes a set Ω_{t-1}, which is information on a set of variables dated t-1 and earlier. We can specify Ω_{t-1} in a variety of ways. For example, Lucas assumed that Ω_{t-1} included information on all lagged values of $p_t(z)$ and lagged values of $y_t(z)$ in all markets. One could equally well conceive of less comprehensive definitions of Ω_{t-1}. As we shall see, the definition of Ω_{t-1} has some important consequences. For now, along with Lucas we suppose that Ω_{t-1} includes a comprehensive list of variables including lagged outputs and prices in all markets. We don't index Ω_{t-1} by z, since all markets are assumed to share the information in Ω_{t-1}.

Thus we have $I_t(z) = (\Omega_{t-1}, p_t(z))$. We will find it convenient to use the recursive projection formula in getting an expression for $P[p_t|I_t(z)]$. By way of doing this, we first obtain the decomposition

(2) $$p_t = P[p_t|\Omega_{t-1}] + \xi_t$$

where ξ_t is a random variable (a least squares disturbance) that by the orthogonality principle obeys $E(\xi_t \cdot P[p_t|\Omega_{t-1}]) = 0$ and $E\xi_t = 0$. Let the variance of ξ_t be denoted $E\xi_t^2 = \sigma^2$.

Next, suppose that demand is distributed so that

(3) $$p_t(z) = p_t + z_t$$

where $Ez_tp_t = 0$, and $Ez_t\xi_t = 0$. We assume that z_t is orthogonal to all variables in Ω_{t-1}. Equation (3) expresses $p_t(z)$ as the sum of the economy-wide price p_t and a random term z_t, which is uncorrelated with p_t and previous information and which measures relative price movements. We assume that $Ez_t = 0$ and $Ez_t^2 = \tau^2$. The specification $Ez_t = 0$ merely means that relative price movements average out across markets.

Substituting (2) into (3) gives

(4) $$p_t(z) = P[p_t|\Omega_{t-1}] + \xi_t + z_t$$

where again $E\xi_tz_t = 0$, so that $E\{(\xi_t+z_t)^2\} = \sigma^2 + \tau^2$. Now what is the linear least squares forecast of $p_t(z)$ given Ω_{t-1}? Since by construction ξ_t is orthogonal to Ω_{t-1}, and since by assumption z_t is orthogonal to Ω_{t-1}, we have that (4) is a projection equation with disturbance $\xi_t + z_t$, so that

(5) $$P[p_t(z)|\Omega_{t-1}] = P[p_t|\Omega_{t-1}].$$

We are now in a position to apply the recursive projection formula to get an expression for $P[p_t | I_t(z)]$:

(6) $$P[p_t|\Omega_{t-1}, p_t(z)] = P[p_t|\Omega_{t-1}]$$

$$+ P[(p_t - P[p_t|\Omega_{t-1}]) | (p_t(z) - P[p_t(z)|\Omega_{t-1}])].$$

Using (2), (4), and (5) we know that

$$p_t - P[p_t|\Omega_{t-1}] = \xi_t$$

$$p_t(z) - P[p_t(z)|\Omega_{t-1}] = \xi_t + z_t$$

and also that $E\xi_t z_t = 0$. The projection of $p_t - P[p_t|\Omega_{t-1}]$ on $p_t(z) - P[p_t(z)|\Omega_{t-1}]$ is therefore given by

$$\phi(p_t(z) - P[p_t(z)|\Omega_{t-1}])$$

where the least squares coefficient ϕ is given by

$$\phi = \frac{E\xi_t \cdot (\xi_t + z_t)}{E[(\xi_t + z_t)^2]}$$

$$= \frac{\sigma^2}{\sigma^2 + \tau^2} .$$

Using the above and recalling that $P[p_t(z)|\Omega_{t-1})] = P[p_t|\Omega_{t-1}]$, we can write (6) as

$$P[p_t|I_t(z)] = P[p_t|\Omega_{t-1}] + \frac{\sigma^2}{\sigma^2 + \tau^2} [p_t(z) - P[p_t|\Omega_{t-1}]]$$

or

(7) $$P[p_t|I_t(z)] = \theta P[p_t|\Omega_{t-1}] + (1-\theta) p_t(z)$$

where $\theta = \frac{\tau^2}{\tau^2 + \sigma^2}$ and $(1-\theta) = \frac{\sigma^2}{\tau^2 + \sigma^2}$.

The parameter θ is the fraction of the conditional variance in $p_t(z)$ due to relative price variation. The larger is this fraction, the smaller is the weight placed on $p_t(z)$ in revising $P[p_t|\Omega_{t-1}]$ to form $P[p_t|I_t(z)]$. This makes sense, since the larger is θ, the more likely is it that a change in $p_t(z)$ reflects a relative rather than a general price change.

Substituting (7) into (1) gives

$$y_t(z) = \gamma[p_t(z)-\theta P[p_t|\Omega_t]-(1-\theta)p_t(z)] + \lambda y_{t-1}(z)$$

or

$$(8) \qquad y_t(z) = \gamma\theta(p_t(z)-P[p_t|\Omega_{t-1}]) + \lambda y_{t-1}(z).$$

Let $g(z)$ be the probability density function of z. An index of average output is then given by

$$y_t = \int y_t(z)g(z)dz$$

which is simply the mean of the distribution of $y_t(z)$. The average price level is given by

$$p_t = \int p_t(z)g(z)dz.$$

Integrating both sides of (8) with respect to $g(z)dz$ (i.e., averaging (8) over all markets) gives

$$\int y_t(z)g(z)dz = \gamma\theta(\int p_t(z)g(z)dz-P[p_t|\Omega_{t-1}]\int g(z)dz)$$

$$+ \lambda\int y_{t-1}(z)g(z)dz$$

or

$$(9) \qquad y_t = \gamma\theta(p_t-P[p_t|\Omega_{t-1}]) + \lambda y_{t-1}.$$

Since $\gamma > 0$ and $0 < \theta = \frac{\tau^2}{\sigma^2+\tau^2} < 1$, equation (9) is a version of a Phillips curve relating output directly to the gap between the average price level p_t and agents' prior forecast of the price level $P[p_t|\Omega_{t-1}]$. To write (9) in an alternative way, solve (9) for $p_t - P[p_t|\Omega_{t-1}]$ to get

$$p_t - P[p_t|\Omega_{t-1}] = (\gamma\theta)^{-1}y_t - \lambda(\gamma\theta)^{-1}y_{t-1}.$$

Adding and subtracting p_{t-1} gives

$$(10) \qquad p_t - p_{t-1} = (\gamma\theta)^{-1}y_t - \lambda(\gamma\theta)^{-1}y_{t-1}$$
$$+ (P[p_t|\Omega_{t-1}]-p_{t-1}),$$

which is in the form of a standard natural rate Phillips curve relating inflation (p_t-p_{t-1}) directly to output and to expected inflation $P[p_t|\Omega_{t-1}] - p_{t-1}$. According to (10), the Phillips curve shifts up in the (p_t-p_{t-1}, y_t) plane by the exact amount of any increase in expected inflation $P[p_t|\Omega_{t-1}] - p_{t-1}$. This characteristic of (10) is often taken as the hallmark of the natural unemployment rate hypothesis. It seems to offer an explanation for why the Phillips curve tradeoff has worsened as average inflation rates have increased over the last decade in many western countries.

It is important to note that the slope parameter $\gamma\theta$ of (9) depends on the ratio of the variances of the random terms ξ_t and z_t, since $\theta = \frac{\tau^2}{(\sigma^2+\tau^2)}$. The larger is the variance of $z_t (Ez_t^2=\tau^2)$, relative to the variance of $\xi_t (E\xi_t^2=\sigma^2)$, the larger is θ and, therefore, the larger is $\theta\gamma$. So the more variable is z_t relative to ξ_t, the larger is the reponse of aggregate output y_t to unexpected aggregate price changes $(p_t - P[p_t|\Omega_{t-1}])$. That is, the larger is the variance of z_t relative to that of ξ_t, the greater is the tendency of rational agents to view a given unexpected increase in price as a relative price change to which their output decision should respond.

An implication of the dependence of the slope of (9) on the ratio of variances of relative to aggregate price movements is that (9) is not predicted to remain unchanged across different aggregate demand regimes. That is, a "favorable" tradeoff between output and unexpected inflation (that is, a large value of $\gamma\theta$) will exist only when σ^2 is small relative to τ^2. An attempt by the authorities to exploit the tradeoff more fully by changing aggregate demand regimes can be expected to increase the variance σ^2 relative to τ^2, and thus change the slope $\gamma\theta$. This is yet another example of how agents' optimal decision rules change in response to changes in the random processes governing the exogenous variables they base their decisions on.

For empirical support of his model, Lucas pointed to evidence that the slope parameter in (9) does indeed seem to be much smaller in regimes with very high variance in nominal aggregate demand than in regimes with low variance of nominal aggregate demand.

Persistence in Output

It is a fact that the unemployment rate and deviations of real GNP from its trend are highly serially correlated, i.e., strongly and positively correlated with their own lagged values. How can this fact be accounted for within the context of the preceding model? The answer depends delicately on how we specify the information set Ω_{t-1}. To begin, suppose that Ω_{t-1} includes enough information for agents to be able to form lagged values of the aggregate price index p_{t-1}, p_{t-2}, ... This could happen if lagged values of the price index were published, or if agents received with a one-period lag prices in all other markets so that they could form p_{t-1}, p_{t-2}, ... for themselves. Further, it is obviously not restrictive at all to assume that agents' information includes their own lagged forecasts $P[p_{t-1}|\Omega_{t-2}]$, $P[p_{t-2}|\Omega_{t-3}]$, ... Since both lagged forecasts and lagged prices are assumed to be included in Ω_{t-1}, it follows that the lagged forecast errors

$$p_{t-1} - P[p_{t-1}|\Omega_{t-2}],\ p_{t-2} - P[p_{t-2}|\Omega_{t-3}],\ \ldots$$

are included in Ω_{t-1}. The least squares orthogonality condition then implies

$$(11) \qquad E\{(p_t-P[p_t|\Omega_{t-1}])(p_{t-j}-P[p_{t-j}|\Omega_{t-j-1}])\} = 0$$

for all $j \geq 1$, since $(p_{t-j}-P[p_{t-j}|\Omega_{t-j-1}])$ is included in Ω_{t-1} (remember the orthogonality principle: the least squares projection $P[p_t|\Omega_{t-1}]$ is uniquely determined by the condition that the forecast error be orthogonal to all components of Ω_{t-1}). According to equation (11), the forecast error is uncorrelated with its own lagged values, i.e., it is serially uncorrelated. This is a direct implication of our having assumed that lagged p's are included in Ω_{t-1}.

In the context of (9), (11) has the implication that if $\lambda = 0$, y_t itself will be serially uncorrelated, since then y_t equals a scalar $\gamma\theta$ times the serially uncorrelated forecast error $p_t - P[p_t|\Omega_{t-1}]$. However, if $\lambda > 0$, the effects of forecast errors which are themselves serially uncorrelated, will persist and make y_t serially correlated.

Since it is a fact that y_t is strongly serially correlated, it is necessary to permit τ to exceed zero (and by a healthy amount) if the facts are to be accounted for in the context of (9). An objection which has been made to this procedure is that it is ad hoc and not derived from any explicitly stated theory. The theoretical content of (9) is entirely reflected in the relative price parameter γ and the signal extraction parameter θ. Conceivably, a cost-of-adjustment model could be cooked up to rationalize the presence of lagged y's in (9). But until such a model is produced, the preceding criticism is well taken.

An alternative way of explaining serial correlation in output while retaining (9) is to relax the assumption that Ω_{t-1} includes lagged aggregate prices. If Ω_{t-1} doesn't include lagged p's, the orthogonality condition no longer implies (11), so that the forecast errors can themselves be serially correlated, and so can account for serially correlated output even with $\lambda = 0$ in (9). One way to think of having Ω_{t-1} failing to include lagged p's is by supposing that the price indexes appropriate to agents' decisions are never collected, so that the published price indexes are error-ridden. Another device, which has been implemented by Lucas,[2/] is to interpret p_t in all of the proceeding as "nominal aggregate demand" rather than price, which, as above, is composed of relative and aggregate movements. If agents never observe nominal aggregate demand, but know its second movements so that they can calculate the least squares projections studied above, serially correlated output can be accounted for within the context of a version of (11).

Footnotes

1/ Robert E. Lucas, Jr., "Some International Evidence on Output-Inflation Tradeoffs," American Economic Review, (June 1973).

2/ See Robert E. Lucas, Jr., "An Equilibrium Model of the Business Cycle," J.P.E., (December 1975).

Investment Under Uncertainty

These pages describe the investment problem faced by a firm operating under uncertainty in a very simple setting. We impose increasing costs of investing at higher absolute rates, which gives rise to a Keynesian investment schedule. The advantages in moving to an explicitly stochastic setting are greater realism and an emphasizing of the extent to which the parameters of the firm's investment schedule will themselves change in response to perceived changes in the random processes governing the variables exogenous to the firm. As Lucas has emphasized, it is critical to recognize this dependence in carrying out econometric policy evaluations.

The setup here is a poor man's version of the one in Lucas and Prescott's important paper on investment.* Our exposition of Bellman's "principle of optimality" is heuristic, to say the least, and is intended only to whet the reader's appetite. The reader is referred to Lucas and Prescott's article for a careful treatment.

The firm produces subject to the production function

$$q_t = k_t$$

where q_t is output in period t and k_t is the capital stock at time t. The firm's capital is linked to its investment by

$$k_{t+1} = k_t + I_t ,$$

where for simplicity we assume no depreciation. The firm sells all the output that it wants in a competitive market at the fixed price p_t. The

---

*Robert E. Lucas, Jr., and Edward C. Prescott, "Investment Under Uncertainty," Econometrica, Vol. 39, No. 5, September 1971.

price p_t is assumed to be governed by a first-order Markov process, so that

(1) $$E[p_{t+1}|p_t, p_{t-1}, \ldots] = \lambda p_t \ , \quad 0 < \lambda$$

and

$$E[p_{t+j}|p_t, p_{t-1}, \ldots] = \lambda^j p_t \ ,$$

where $E[Y|X]$ denotes the mathematical expectation of Y given X. We assume that $p_t > 0$ and $E[p_{t+j}|p_t] > 0$ for all t and j. The firm's revenue at time t is $p_t q_t = p_t k_t$.

Investment at time t is assumed to take place subject to the cost of investment function

$$h_t I_t + \frac{1}{2} J_t I_t^2$$

where h_t, $J_t > 0$ measure the cost of acquiring new investment goods. Notice that the cost of adjustment is quadratic in investment and so satisfies the restrictions on the cost of adjustment schedule that we earlier imposed in order to derive a Keynesian investment schedule in a deterministic framework.

The firm chooses investment to maximize its expected present value

(2) $$v_t = E_t \sum_{j=t}^{\infty} \beta^{j-t} [p_j k_j - h_j I_j - \frac{1}{2} J_j I_j^2]$$

subject to

(3) $$k_j = k_{j-1} + I_{j-1}$$

k_t fixed from the past.

Here E_t is the mathematical expectation operator conditioned on the firm's information as of time t. We assume that the firm's information consists of current and past values of p, h, and J and, of course, its current capital stock k_j. The firm is assumed to know the conditional probability distributions that govern how subsequent p, h, and J's evolve from current and past ones, i.e., the firm knows the parameters of the Markov process (1), and also knows the analogous processes for h and J. The parameter β is a discount factor that can be interpreted as equal to the reciprocal of one plus the nominal rate of interest.

Substituting (3) into (2), we can write the firm's expected present value as

$$v_t = (p_t k_t - h_t I_t - \frac{1}{2} J_t I_t^2) + \beta E_t[(k_t + I_t)p_{t+1} - h_{t+1} I_{t+1} - \frac{1}{2} J_{t+1} I_{t+1}^2]$$
$$+ \beta^2 E_t[(k_t + I_t + I_{t+1})p_{t+2} - h_{t+2} I_{t+2} - \frac{1}{2} J_{t+2} I_{t+2}^2]$$
$$+ \ldots$$

From (1), we have $E_t p_{t+j} = \lambda^j p_t$. Substituting this into the preceding equation gives

$$v_t = (p_t k_t - h_t I_t - \frac{1}{2} J_t I_t^2) + \beta[(k_t + I_t)\lambda p_t - E_t(h_{t+1} I_{t+1} - \frac{1}{2} J_{t+1} I_{t+1}^2)]$$
$$+ \beta^2[(k_t + I_t + I_{t+1})\lambda^2 p_t - E_t(h_{t+2} I_{t+2} - \frac{1}{2} J_{t+2} I_{t+2}^2)]$$
$$+ \ldots$$

or

$$(4) \qquad v_t = (p_t k_t - h_t I_t - \frac{1}{2} J_t I_t^2)$$
$$+ p_t \sum_{j=t+1}^{\infty} \beta^{j-t} \lambda^{j-t} (k_t + I_t + \ldots + I_{j-1})$$
$$- E_t \sum_{j=t+1}^{\infty} \beta^{j-t} (h_{t+j} I_{t+j} - \frac{1}{2} J_{t+j} I_{t+j}^2) .$$

The firm maximizes v_t with respect to current investment I_t. Differentiating v_t in (4) with respect to I_t and equating the result to zero gives

$$\frac{\partial v_t}{\partial I_t} = -h_t - J_t I_t + p_t \sum_{j=t+1}^{\infty} \beta^{j-t} \lambda^{j-t} = 0$$

or

$$h_t + J_t I_t = p_t \left(\frac{\beta\lambda}{1-\beta\lambda} \right) .$$

The firm thus equates the marginal cost of investment at time t, $h_t + J_t I_t$, to the expected discounted revenue, which equals $\beta\lambda p_t/(1-\beta\lambda)$. Solving for I_t gives

$$(5) \qquad I_t = - \frac{h_t}{J_t} + \frac{p_t}{J_t} \left(\frac{\beta\lambda}{1-\beta\lambda} \right) .$$

Equation (5) is in the form of a simple Keynesian investment schedule. Investment I_t varies directly with the product price p_t and inversely with the parameter J_t that helps to govern the costliness of investing rapidly. Notice the dependence of investment on both the discount factor β and the Markov parameter λ. The presence of λ in (5) shows that the form of the investment schedule (5) is itself dependent on the stochastic process governing the variables exogenous to the firm, in this case p_t. The investment schedule (5) is therefore <u>not</u> predicted to be invariant with respect to an intervention that changes the stochastic process for $\{p_t\}$.

Actually, one would on average expect there to be a relation between λ and β, since λ governs the expected rate of inflation in the product price. That is,

$$E_t(p_{t+1}-p_t) = \lambda p_t - p_t$$

$$= (\lambda-1)p_t ,$$

so that $(\lambda-1)$ is the expected rate of inflation of commodity prices for this firm. Let us write β as

$$\beta = \frac{1}{1+\rho+\pi}$$

where ρ is the real rate of interest facing the firm and $\pi = (\lambda-1)$ is the expected rate of inflation in output prices. Then we can write

$$\beta \cdot \lambda = \frac{\lambda}{1+\rho+(\lambda-1)} = \frac{\lambda}{\rho+\lambda} .$$

It follows that

$$\frac{\beta\lambda}{1-\beta\lambda} = \frac{\lambda}{\rho} .$$

Therefore (5) can be written in the form

(6) $$I_t = -\frac{h_t}{J_t} + \frac{p_t}{J_t} \cdot \frac{\lambda}{\rho} .$$

This is in the form of a Keynesian investment schedule in which investment varies directly with output price p_t and inversely with the cost of investment parameter J_t and the real interest rate ρ.

For expository purposes, we have specified the technology and the market structure in the preceding example so that an explicit investment function could be calculated using simple techniques. For this simplicity, it is essential in (4) that the current investment decision can optimally be determined independently of future investment decisions. This allows a single marginal condition to determine the investment schedule. The assumption that the firm is a perfect competitor in the output market is what enabled us to proceed in this way. It is instructive to study the investment decision under more general conditions in which the preceding simplifications are absent.

For example, let us assume the same setup as above, except that the firm faces a downward sloping demand schedule so that its price obeys

$$p_t = p(q_t, u_t) = p(k_t, u_t)$$

where u_t is a random demand-shift term that follows a first-order Markov process. When the firm stands at time 0, its objective is to maximize

$$\text{(7)} \qquad v_0 = k_0 p(k_0, u_0) - h_0 I_0 - \frac{1}{2} J_0 I_0{}^2 + E_0 \sum_{j=1}^{\infty} \beta^j [k_j p(k_j, u_j) - h_j I_j - \frac{1}{2} J_j I_j{}^2]$$

subject to

$$k_j = k_{j-1} + I_{j-1} \qquad j \geq 1 .$$

k_0 given.

Proceeding as before, suppose we differentiate v_0 with respect to I_0 to obtain

$$\text{(8)} \qquad \frac{\partial v_0}{\partial I_0} = -h_0 - \frac{1}{2} J_0 I_0 + \sum_{j=1}^{\infty} \beta^j E_0 \{p_j + k_j \frac{\partial p_j}{\partial k_j}\} = 0.$$

Now $k_j = k_0 + I_0 + I_1 + \ldots + I_{j-1}$, so that the term under the sum is a function of the random variables $I_1, \ldots, I_j, \ldots$, which are yet to be determined. Thus, (8) is one equation in $I_0, I_1, \ldots$ and by itself is incapable of determining I_0. (Previously, we assumed that $\partial p_j/\partial k_j = 0$, which caused (8) to collapse to one equation in the unknown I_0.)

To proceed, suppose we differentiate (7) with respect to I_1 and set the result to zero to obtain

$$(9) \qquad \frac{\partial v_0}{\partial I_1} = \beta E_0\{-h_1 - \frac{1}{2} J_1 I_1 + \sum_{j=2}^{\infty} \beta^{j-1} E_1(p_j + \frac{\partial p_j}{\partial k_j} \cdot k_j)\} = 0$$

where we have used the fact that $E_0 X_j = E_0(E_1 X_j)$ to write (9). Now the term in braces is a random variable at time 0, a random variable of which the firm is instructed by (9) to take the expected value. But notice that the term in braces is simply the right side of (8) shifted forward one period. Equation (9) has the natural interpretation of instructing the firm to expect as of time 0 that it will do the right thing with respect to I_1 when time 1 rolls around; namely, set the term in braces equal to zero as (8) instructs. Thus, (9) instructs the firm to expect that it will do the optimal thing next period given next period's information, and to calculate this period's best action on that assumption. This is Bellman's "principle of optimality." Proceeding in a similar way by differentiating v_0 with respect to $I_2, I_3, \ldots$, a sequence of marginal conditions analagous to (9) can be derived. Heuristically, this set of equations together with (8) can be viewed as determining I_0 jointly with forecasts as of time 0 of future investment rates I_j, $j \geq 1$.

Proceeding more formally, notice that v_0 in (7) can be written as

(10) $$v_0 = k_0 p(k_0,u_0) - h_0 I_0 - \frac{1}{2} J_0 I_0^{\ 2} + \beta E_0 v_1$$

where $v_1 = E_1 \sum_{j=1}^{\infty} \beta^{j-1} [p_j k_j - h_j I_j - \frac{1}{2} J_j I_j^{\ 2}]$.

Let u_t, h_t, and J_t be governed by first-order Markov processes so that at time t all pertinent information is summarized by the information set (u_t, h_t, J_t, k_t), which completely characterizes the state of the firm, that is, its present position and likely future position insofar as the present and past help forecast the future. Let $v(u_0,h_0,J_0,k_0)$ denote the maximum of (10) with respect to I_0. The same function v( ) evaluated at (u_1,h_1,J_1,k_1) must denote the maximum of present value v_1 at time 1 with respect to I_1. Thus, a maximum of (10) with respect to I_0 must satisfy

(11) $$v(u_0,h_0,J_0,k_0) = \max_{I_0} \{k_0 p(k_0,u_0) - h_0 I_0 - \frac{1}{2} J_0 I_0^{\ 2} + \beta E_0 v(u_1,h_1,J_1,k_0+I_0)\} .$$

Equation (11) is a functional equation to be solved for the valuation function v(u,h,J,k), which tells the maximum present value of the firm as a function of the variables u, h, J, and k that characterize its present state. Under suitable conditions, there exists a unique solution to this functional equation, the solution being a function v(u,h,J,k). Once such a solution is obtained, it is possible to characterize the investment schedule. To illustrate, suppose that v(u,h,J,k) turns out to be continuous, concave, and differentiable. Then differentiate the right side of (11) to obtain the marginal condition

(12) $$-h_0 - J_0 I_0 + \beta E_0 \frac{\partial v}{\partial k} (u_1,h_1,J_1,k_0+I_0) = 0 .$$

The solution of this equation is the investment schedule. As indicated earlier in our simple example, the parameters of the Markov processes governing u_t, h_t, and J_t appear in the investment schedule by virtue of the term $E_0 \frac{\partial v}{\partial k}(\ )$, since this conditional expectation inherits its form partly from those Markov processes. The form of the decision rule (12) is thus a function of the stochastic processes that agents perceive to be governing the exogenous variables they face. Interventions that alter those perceptions cannot be expected to leave unaltered the forms of the decision rules.

Exercise:

i. Derive the investment schedule in the case where the firm is a perfect competitor in the output market, and in which output prices follow the second-order Markov process

$$E[p_{t+1}|p_t,p_{t-1},p_{t-2}, \ldots] = \lambda_1 p_t + \lambda_2 p_{t-1}$$

$$E[p_{t+2}|p_t,p_{t-1}] = \lambda_1 E[p_{t+1}|p_t,p_{t-1}] + \lambda_2 p_t$$

$$E[p_{t+j}|p_t,p_{t-1}] = \lambda_1 E[p_{t+j-1}|p_t,p_{t-1}] + \lambda_2 E[p_{t+j-2}|p_t,p_{t-1}]$$

for $j \geq 3$.

Optimal Monetary Policy

The central practical issue separating Keynesian from non-Keynesian economists is the nature of the optimal feedback rules for setting monetary and fiscal policy instruments. Keynesian economists have advocated "activist" policies, which incorporate feedback from current and past observations on the state of the economy to future settings of fiscal and monetary instruments (e.g., the deficit and the money supply). Usually, these feedback rules are thought to imply that policy ought to "lean against the wind," calling for increases in taxes and lower rates of growth in the money supply in the boom, and lower taxes and higher growth in money when a recession is in the offing.

On the other hand, non-Keynesian economists such as Henry Simons and Milton Friedman have advocated that the government follow rules without feedback in setting fiscal and monetary policy. In essence, Simons and Friedman's advice to the government is three-fold. First, set government expenditures on the basis of cost-benefit considerations and don't manipulate government expenditures to try to combat the business cycle. Second, keep tax rates fixed at levels that, given the rate of government expenditures, make the rate of growth of government debt average out over the business cycle to some desired level. Third, make the money supply grow at a constant rate of x percent per year, regardless of the state of business conditions. The rate x should be set with a view to the average rate of inflation desired.

The differences between the prescriptions of the two schools do not seem to be attributable to any differences over the goals each of them would like policy to achieve. Each would like to keep the economy as close as possible to "full employment," hopefully with a stable price level. Neither can the difference between policy prescriptions be traced to any differences over how to derive a feedback rule for a given economic model and a given objective function: that is essentially a technical matter about which there is no room for disagreement. Rather, the disagreement stems from fundamental differences on the question of what is the correct macroeconomic model. In particular, a great deal hinges on the question of how to model the manner in which agents form their expectations about future events.

These pages first state the case for using rules with feedback. The argument for using rules with feedback assumes that the economy can be described as a set of stochastic difference equations (i.e., an econometric model), which, when written in a particular form, has coefficients that are invariant across alternative feedback rules that the authority might use. Given this setup, rules with feedback will be shown to dominate rules without feedback. This setup displays the intellectual foundations of the Keynesian "activist" policy strategy.

We next explore the parts of the preceding setup that a non-Keynesian economist might question. In particular, it will be seen that the form of the model that must remain invariant across changes in feedback rules is one that embodies the public's rule for forecasting future prices. As we have seen repeatedly, if agents optimize, the forms of such forecasting rules depend on the nature of the exogenous stochastic processes facing them. Since changes in the government's feedback rules alter those processes, the forecasting rules and therefore the parameters of the model will change with each change in the policy feedback rule. We shall see that a defense of the Simons-Friedman rules without feedback can be erected on the foundations of this observation.

Optimal Control with *Ad Hoc* Expectations

Suppose that the economy is governed by the following simple macroeconomic model:

(1) $$y_t = \gamma(p_t - {}_tp_{t-1}^*) + \lambda y_{t-1} + u_t \qquad \gamma > 0$$

(2) $$m_t - p_t = y_t + \varepsilon_t$$

(3) $${}_tp_{t-1}^* = v(L)p_{t-1} (\equiv \sum_{i=0}^{\infty} v_i p_{t-i-1})$$

where y_t = log of real GNP, p_t = log of the GNP deflator, m = log of the money supply, and ${}_tp_{t-1}^*$ is the public's expectation of the log of the price level at time t, the expectation being formed at time t-1. We assume that u_t and ε_t are each serially independent, stationary random processes with means of zero and finite variances. For simplicity, we assume that ε and u are uncorrelated, so that $E\, u_t\varepsilon_t = 0$. Equation (1) is a simple Phillips curve embodying the natural unemployment rate hypothesis, since only unexpected increases in the price level are posited to boost aggregate supply. Equation (2) is a simple portfolio balance schedule that excludes the interest rate, for simplicity only. Equation (3) describes how expectations are formed as a weighted sum of past prices. The model (1), (2), and (3) determines stochastic processes for y_t, p_t and ${}_tp_{t-1}^*$ as functions of the disturbance processes u_t, ε_t, and the money supply process m_t.

The goal of the monetary authority is to choose a stochastic process for the money supply that in some sense optimizes the performance of the economy. To illustrate this problem in the simplest context, we assume initially that the object of the monetary authority is to minimize the mean squared error of real GNP around some fixed target level y*,

$$(4) \qquad M.S.E. = E(y_t - y^*)^2 .$$

As we will indicate in more detail later, this criterion is a sensible one for comparing rules that have been in effect for a long time, or else in comparing the eventual performances of alternative rules after they will have been in effect for a long time. It is convenient to decompose the mean squared error as follows:

$$E[(y_t - y^*)^2] = E[((y_t - Ey_t) + (Ey_t - y^*))^2]$$

$$= E((y_t - Ey_t))^2 + E((Ey_t - y^*)^2)$$

$$+ 2E(y_t - Ey_t)(Ey_t - y^*).$$

Since $Ey_t - y^*$ is not random, and since $E(y_t - Ey) = Ey_t - Ey_t = 0$, we have the decomposition

$$(5) \qquad M.S.E. = E[(y_t - Ey_t)^2] + (Ey_t - y^*)^2 ,$$

which expresses the mean squared error as the sum of the variance of y and the "bias squared" around y^*.

To derive an optimal monetary policy rule, it is convenient first to solve for a "final form" for y_t expressing y_t as a function of current and lagged u's, ε's, and m's. Substituting (3) into (1) gives

$$(1 - \lambda L) y_t = \gamma(1 - Lv(L)) p_t + u_t .$$

Solving (2) for p_t and substituting into the above equation gives

$$(1 - \lambda L) y_t = \gamma[1 - Lv(L)](m_t - y_t - \varepsilon_t) + u_t$$

or

$$[1 + \gamma - (\lambda + \gamma v(L))L] y_t = \gamma[1 - Lv(L)](m_t - \varepsilon_t) + u_t .$$

Assuming that $[1+\gamma-(\lambda+\gamma v(L))L]$ has a stable inverse that is one-sided in nonnegative powers of L, we have

$$y_t = \frac{\gamma(1-Lv(L))}{(1+\gamma)-(\lambda+\gamma v(L))L} (m_t - \varepsilon_t) + \frac{1}{(1+\gamma)-(\lambda+\gamma v(L))L} u_t$$

or

(6) $$y_t = H(L)m_t + G(L)\varepsilon_t + F(L)u_t$$

where

$$H(L) = \frac{\gamma(1-Lv(L))}{(1+\gamma-(\lambda+\gamma v(L))L} = \sum_{i=0}^{\infty} h_i L^i$$

$$G(L) = -H(L) = \sum_{i=0}^{\infty} g_i L^i$$

$$F(L) = \frac{1}{1+\gamma-(\lambda+\gamma v(L))L} = \sum_{i=0}^{\infty} f_i L^i .$$

Since H(L), G(L), and F(L) are each one-sided, (6) is equivalent with

(6') $$y_t = \sum_{i=0}^{\infty} h_i m_{t-i} + \sum_{i=0}^{\infty} g_i \varepsilon_{t-i} + \sum_{i=0}^{\infty} f_i u_{t-i} .$$

Equation (6) is the final form for y_t.

The monetary authority is assumed to consider linear feedback rules of the form[1/]

(7) $$m_t = k + A(L)\varepsilon_{t-1} + \beta(L)u_{t-1}$$

where $$A(L) = \sum_{i=0}^{\infty} a_i L^i$$

$$\beta(L) = \sum_{i=0}^{\infty} b_i L^i .$$

Through (7), the monetary authority permits itself to react to unexpected changes in the position of the economy, which are completely described by the disturbances u_t and ε_t. The authority is assumed to know the structure of the economy (it knows the parameters of (1), (2), (3), and therefore of (6)). It also has data in all lagged values of y_t, p_t, and m_t, so that it can calculate the lagged disturbances u_t and ε_t by using its knowledge of the parameters of (1), (2), and (3), together with its data on p_t, y_t, and m_t. The monetary authority is required to set m_t before it receives information on current y_t and p_t, which explains why it is permitted only to feedback on lagged ε's and u's in (7).

The goal of the monetary authority is to choose k, A(L), and $\beta(L)$ to minimize the mean squared error (5), subject to the structure of the economy (6). The authority assumes that the parameter values in (6) remain unchanged as it contemplates the effects of alternative choices of k, the a's, and the b's in (7).

To determine the effects of operating under the feedback rule (7), substitute (7) into (6) to obtain

$$y_t = H(L)[A(L)L\varepsilon_t+\beta(L)Lu_t+k] + G(L)\varepsilon_t + F(L)u_t$$

$$y_t = kH(1) + [H(L)A(L)L+G(L)]\varepsilon_t + [H(L)\beta(L)L+F(L)]u_t \; . \tag{8}$$

By substituting (7) into (6), we in effect assume that (7) has been operating forever. For convenience let

$$H(L)A(L)L = \phi(L) = \sum_{i=1}^{\infty} \phi_i L^i$$

$$H(L)\beta(L)L = \psi(L) = \sum_{i=1}^{\infty} \psi_i L^i \; .$$

Since $E\varepsilon_t = Eu_t = 0$ for all t, we have that

$$Ey_t = kH(1) \; .$$

The bias squared term in (5) is minimized by choosing k so that $Ey_t = y^*$:

$$Ey_t = y^* = kH(1) .$$

This is accomplished by setting $k = y^*/H(1)$. Since the variance of y turns out not to depend on k, the mean square error is minimized by minimizing the variance and bias squared separately.

From (8), the variance of y_t is given by

$$\begin{aligned} \text{var } y_t = \text{var } \varepsilon_t \, [g_0^2 + \sum_{j=1}^{\infty} (g_j+\phi_j)^2] \\ + \text{var } u_t \, [f_0^2 + \sum_{j=1}^{\infty} (f_j+\psi_j)^2] . \end{aligned}$$

To find the minimizing values of ϕ_j, for example, we require

$$\frac{\partial \text{ var } y_t}{\partial \phi_i} = \text{var } \varepsilon_t \cdot 2(g_i + \phi_i) = 0$$

so that we set $\phi_i = -g_i$. Similarly, we set $\psi_j = -f_j$. It is readily verified that the second-order conditions for a minimum are satisfied. Thus, we are instructed to set $\phi(L)$ and $\psi(L)$ so that

$$\phi(L) = H(L)A(L)L = -\sum_{i=1}^{\infty} g_i L^i$$

$$\psi(L) = H(L)\beta(L)L = -\sum_{i=1}^{\infty} f_i L^i .$$

Solving for A(L) and $\beta(L)$ gives

$$A(L) = \frac{-\sum_{i=1}^{\infty} g_i L^i}{H(L)L}$$

$$(9) \qquad \beta(L) = \frac{-\sum_{i=1}^{\infty} f_i L^i}{H(L)L}$$

$$k = y^*/H(1) .$$

Choosing A(L), β(L), and k according to (9) minimizes the mean squared error of y_t around y*. Notice that under the rule, we have that $H(L)A(L)L + G(L) = g_0$, $H(L)\beta(L)L + F(L) = f_0$ so that (8) becomes

$$(10) \qquad y_t = H(1)k + g_0\varepsilon_t + f_0u_t$$

$$= y^* + g_0\varepsilon_t + f_0u_t \, .$$

Under the optimal rule (9), once it has been in effect for a long time, y equals y* plus an irreducible, serially uncorrelated noise $(f_0u_t+g_0\varepsilon_t)$. The mean squared error under the rule equals $g_0^2 \text{ var } \varepsilon_t + f_0^2 \text{ var } u_t$. Notice that under the optimal rule, all serial correlation in y has been eliminated. Strong positive serial correlation in y is what most economists mean when they refer to the "business cycle." Under the optimal rule, the business cycle, then, has been eradicated.

The fact that there exists a rule that eliminates serial correlation in output underlies the Keynesian economist's practice of assigning the blame for recessions to inappropriate monetary and fiscal policy. Given the kind of setup employed here, that practice is entirely justified.

It is useful to work out the optimal monetary rule under slightly different conditions. In particular, suppose that at time t-1 it is desired to set m_t so as to minimize $E_{t-1}(y_t-y^*)^2$ where $E_{t-1}(\cdot)$ is the mathematical expectation operator conditioned on information known at time t-1. The authority desires to minimize $E_{t-1}(y_t-y^*)^2$, taking as given the (possibly very unwise) monetary policy that was pursued in the past. As before, it is possible to decompose the mean squared error:

$$(11) \qquad E_{t-1}(y_t-y^*)^2 = E_{t-1}((y_t-E_{t-1}y_t)^2) + (E_{t-1}y_t-y^*)^2,$$

where the first term is the conditional variance around the conditional mean and the second term is the conditional bias squared.

The final form (6) can be written

(12) $$y_t = h_0 m_t + H_1(L) m_t + g_0 \varepsilon_t + G_1(L)\varepsilon_t + f_0 u_t + F_1(L) u_t$$

where
$$H_1(L) = \sum_{i=1}^{\infty} h_i L^i$$
$$G_1(L) = \sum_{i=1}^{\infty} g_i L^i$$
$$F_1(L) = \sum_{i=1}^{\infty} f_i L^i \ .$$

Since ε_t and u_t are serially independent, we have $E_{t-1}u_t = E_{t-1}\varepsilon_t = 0$. Then using (12), we have

$$E_{t-1}y_t = h_0 E_{t-1}m_t + H_1(L)m_t + G_1(L)\varepsilon_t + F_1(L)u_{t-1} \ .$$

To minimize the bias squared, we equate $E_{t-1}y_t$ to y^* to get

$$E_{t-1}m_t = \frac{1}{h_0} y^* - \frac{H_1(L)}{h_0} m_t - \frac{G_1(L)}{h_0} \varepsilon_t - \frac{F_1(L)}{h_0} u_t \ .$$

Under any rule, we have

$$y_t - E_{t-1}y_t = h_0(m_t - E_{t-1}m_t) + g_0\varepsilon_t + f_0 u_t \ .$$

The variance $E_{t-1}((y_t - E_{t-1}y_t)^2)$ is therefore minimized by setting $m_t = E_{t-1}m_t$. Thus, the optimal rule for setting m_t is

(13) $$m_t = \frac{1}{h_0} y^* - \frac{H_1(L)}{h_0} m_t - \frac{G_1(L)}{h_0} \varepsilon_t - \frac{F_1(L)}{h_0} u_t \ .$$

Notice that under this rule y_t obeys

$$y_t = y^* + g_0\varepsilon_t + f_0 u_t \ ,$$

which is derived by substituting (13) into (12). Thus, the rule succeeds in setting y_t to the target y^* plus the irreducible noise $g_0\varepsilon_t + f_0u_t$.

To find out what happens if this rule is followed for a long time, we have only to solve the difference equation (13) for an m_t sequence that satisfies it for all t. To do this, write (13) as

$$h_0m_t + H_1(L)m_t = y^* - G_1(L)\varepsilon_t - F_1(L)u_t$$

or

$$H(L)m_t = y^* - (\sum_{i=1}^{\infty} g_iL^i)\varepsilon_t - (\sum_{i=1}^{\infty} f_iL^i)u_t .$$

Operating on the above equation by $H(L)^{-1}$ (<u>this</u> is where the assumption that the rule has been operating forever comes in, since we are in effect eliminating lagged m's by substituting appropriately lagged versions of (13) for them) gives

$$m_t = \frac{y^*}{H(1)} - \frac{\sum_{i=1}^{\infty} g_iL^i}{H(L)}\varepsilon_t - \frac{\sum_{i=1}^{\infty} f_iL^i}{H(L)}u_t ,$$

which is exactly the rule given by (9).

It bears repeating that the assumption under which the optimal rule is calculated is that the parameters of the final form will remain invariant when the authority departs from its previous rule and implements any new one. As we shall see, this is exactly the point at which rational expectations theorists would question the relevance of the preceding calculations.

<u>An Example</u>

Suppose that the reduced form for real GNP is

$$y_t = \lambda y_{t-1} + b_0m_t + b_1m_{t-1} + \varepsilon_t$$

where ε_t is a serially independent process with mean zero and finite variance. It follows that $E_{t-1}\varepsilon_t = 0$. We assume that the authority desires to minimize $E_{t-1}(y_t-y^*)^2$. The optimal rule must satisfy

$$E_{t-1}y_t = y^*$$

so that

$$E_{t-1}y_t = \lambda y_{t-1} + b_0 E_{t-1}m_t + b_1 m_{t-1} = y^* .$$

Solving for $E_{t-1}m_t$ gives the optimal rule

$$E_{t-1}m_t = m_t = \frac{1}{b_0} y^* - \frac{\lambda}{b_0} y_{t-1} - \frac{b_1}{b_0} m_{t-1} .$$

The Information Variable Problem

Following Kareken, Muench, and Wallace, consider the following problem. The monetary authority desires to set its instrument m_t to minimize $P_{t-1}((y_t-y^*)^2)$, which is the projection of $(y_t-y^*)^2$ on the authority's information at time t-1. Its information set at time t-1 consists of a set of variables Ω_{t-1}. Since the authority makes policy almost continuously, we should think of increments in t as being very small units of time, e.g., days or hours. For this reason, it becomes important to recognize that the authority receives information about some variables (e.g., interest rates, about which it receives information daily) much more often than other variables (e.g., the average quarterly value of GNP, about which it receives information only quarterly). The monetary authority is assumed to have a model of the economy. The important thing about this model is that it supplies the monetary authority with a complete account of the first and second moments of all variables of interest (whether observed or unobserved) conditional on current and lagged settings of the money supply. Further, the model is supposed to tell the monetary authority the effect of alternative settings of the (nonrandom) variable m_t given past random variables. In effect, then, the model supplies the authority with all of the information that it needs to calculate the linear least squares projection of y_t against m_t and Ω_{t-1}:

$$(14) \qquad P[y_t|\Omega_{t-1},m_t] = \beta m_t + \alpha\Omega_{t-1}$$

where α is a vector conformable with Ω_{t-1}. The authority will minimize $E_{t-1}\{y_t-y^*\}^2$ over the class of linear decision rules by setting

$$P[y_t|\Omega_{t-1},m_t] = y^*$$

or

$$(15) \qquad m_t = \frac{y^*}{\beta} - \frac{\alpha}{\beta}\Omega_{t-1} .$$

Now according to the optimal rule (15), what information should the authority respond to in setting m_t? In general, it should "look at everything." More precisely, it should make m_t feedback upon any component of Ω_{t-1} which bears a nonzero component of α; that is, any component of Ω_{t-1} that helps predict the variable y_t that it is interested in controlling.

An Example

Suppose that the reduced form for real GNP is

$$(16) \qquad y_t = \lambda y_{t-1} + b(L)m_t + c(L)r_{t-1} + \varepsilon_t \qquad 0 < \lambda < 1$$

where ε_t is a serially independent stationary random process that obeys the least squares orthogonality conditions $E\,\varepsilon_t y_{t-s} = 0$ for $s \geq 1$ and $E\,\varepsilon_t m_{t-s} = E\,\varepsilon_t r_{t-s-1} = 0$ for <u>all</u> s. The lag operators $b(L) = \sum_{i=0}^{\infty} b_i L^i$ and $c(L) = \sum_{i=0}^{\infty} c_i L^i$ are one-sided on the present and past. By virtue of the least squares orthogonality conditions, equation (16) is a regression equation. Assume that the object of the monetary authority is to minimize the mean squared error $P_{t-1}((y_t - y^*)^2)$. However, while the authority wants to control y_t, it never receives reliable data on y or any lagged values of y. Instead, its information set Ω_{t-1} consists only of lagged values of m and the interest rate r. So the best that the authority can do is to pursue the rule implied by setting

$$P[y_t | \Omega_{t-1}, m_t] = y^* .$$

To calculate $P[y_t | \Omega_{t-1}, m_t]$ from (16), eliminate lagged y's from (16):

$$(1-\lambda L) y_t = b(L) m_t + c(L) r_{t-1} + \varepsilon_t$$

(17) $$y_t = \frac{b(L)}{1-\lambda L} m_t + \frac{c(L)}{1-\lambda L} r_{t-1} + \frac{1}{1-\lambda L} \varepsilon_t .$$

By virtue of the strong orthogonality conditions imposed on ε, in particular that ε_t be orthogonal to r_s and m_s for <u>all</u> s, it follows that the composite disturbance $\frac{1}{1-\lambda L} \varepsilon_t$ in (17) is orthogonal to m_s and r_s at all lags. Therefore (17) is a regression equation, so that

$$P[y_t | \Omega_{t-1}, m_t] = \frac{b(L)}{1-\lambda L} m_t + \frac{c(L)}{1-\lambda L} r_{t-1} .$$

Assuming that the above projection is invariant with respect to changes in the feedback rule for m, the optimal feedback rule is then

$$\frac{b(L)}{1-\lambda L} m_t + \frac{c(L)}{1-\lambda L} r_{t-1} = y*$$

or

(18) $$b_0 m_t = (1-\lambda)y* - c(L)r_{t-1} - (\sum_{i=1}^{\infty} b_i L^i)m_t .$$

Under the optimal rule (18), output moves according to

(19) $$y_t = \lambda y_{t-1} + (1-\lambda)y* + \varepsilon_t$$

which is derived by substituting (18) into (16). Notice that output is serially correlated under the optimal rule, this being a consequence of the information set's not including observations on lagged y's.

In the context of this example, any information that will help predict y_t ought to be included in the feedback rule determining the money supply.

This example exhibits the logic underlying the Keynesian case for "activist" monetary and fiscal policy. The notion that the economy can be described by presumably a large system of stochastic difference equations with fixed parameters underlies the standard Keynesian objections to the monism of monetarists who argue that the monetary authority should ignore other variables such as interest rates and concentrate on keeping the money supply on a steady growth path. The Keynesian view that, on the contrary, the monetary authority should "look at (and respond to) everything" including interest rates, rests on the following propositions:[2/] (a) The economic structure is characterized by extensive simultaneity, so that shocks that impinge on one variable, e.g., an interest rate, impinge also on most others; (b) Due to lags in the system, the effects of shocks on the endogenous variables are distributed over time, and so are serially correlated and therefore somewhat predictable; and (c) The "structure" of these lags is constant over time and does not depend on how the monetary authority is behaving. These propositions imply that variables that the authority observes very frequently, (e.g., daily, such as interest rates), carry information useful for revising its forecasts of future value of variables that it can't observe as often, such as GNP and unemployment. This follows because the same shocks are affecting both the observed and the unobserved variables, and because those shocks have effects that persist. It follows then from (c) that the monetary authority should in general revise its planned setting for its policy instruments each time it receives some new and surprising reading on a variable that is determined simultaneously with a variable like GNP or unemployment that it is interested in controlling.

Such an argument eschewing a simple x-percent growth rate rule in favor of "looking at everything" has been made by Paul Samuelson [7]:

> ... when I learned that I had been wrong in my beliefs about how fast M was growing from December 1968 to April 1969, this news was just one of twenty interesting items that had come to my knowledge that week. And it only slightly increased my forecast for the strength of aggregate demand at the present time. That was because my forecasts, so to speak, do not involve "action at a distance" but are loose Markov processes in which a broad vector of current variables specify a "phase space" out of which tomorrow's vector develops. (In short, I knowingly commit that most atrocious of sins in the penal code of the monetarists--I pay a great deal of attention to all dimensions of "credit conditions" rather than keeping my eye on the solely important variable $\dot{M}/M$.)

> ... often, I believe, the prudent man or prudent committee can look ahead six months to a year and

> with some confidence predict that the economy will be in other than an average or "ergodic" state. Unless this assertion of mine can be demolished, the case for a fixed growth rate for M, or for confining M to narrow channels around such a rate, melts away.
>
> ... These general presumptions arise out of what we know about plausible models of economics and about the findings of historical experience.[3/]

Optimal Control Under Rational Expectations

The preceding pages provide a simple but complete description of current procedures for macroeconometric policy evaluation. First, a macroeconometric model is estimated. Then, final form equations of the form of equation (6) are derived from the estimated model expressing each of the endogenous variables that enter the authority's objective function as function of the policy instrument and current and lagged exogenous variables and disturbances. Those equations are assumed invariant as changes in the rule are assumed and their effects on the objective function are evaluated.

A critical aspect of the above procedure is the implicit assumption that agents' decision rules, which are impounded, for example, in the estimated investment schedule, the consumption function, and so on, remain unchanged in the face of alternative stochastic processes for the control variable that different feedback rules imply. As we have repeatedly seen, however, optimal decision rules invariably respond to changes in the stochastic processes governing the exogenous variables facing agents. Thus, the invariance assumption needed to validate the preceding case for rules with feedback will not in general hold where agents use optimal decision rules. We propose to illustrate how taking this into account can drastically change the implied optimal control rule.

Return to the model formed by (1), (2), and (3), but replace (3) with the assumption that the public's expectations are rational, so that ${}_tp_{t-1}^* = E\,p_t|\Omega_{t-1}$, where Ω_{t-1} is the public's information set. Making this change has the very important consequence of introducing a dependence of the parameters of the final form upon the parameters of the rule chosen by the authority. As mentioned above, the preceding calculations assumed that there were no such dependences. As will be seen, the fact that assuming rational expectations induces such dependencies has very serious policy implications.

The model under rational expectations becomes

$$(1) \qquad y_t = \gamma(p_t - {}_tp_{t-1}^*) + \lambda y_{t-1} + u_t$$

$$(2) \qquad m_t - p_t = y_t + \varepsilon_t$$

$$(3') \qquad {}_tp_{t-1}^* = E(p_t|\Omega_{t-1})$$

where the only change in the model has been to replace the assumption of fixed autoregressive expectations in (3) with the rational expectations hypothesis embodied in (3'). Let us assume that the public and the monetary authority share the same information set Ω_{t-1}, which consists of (at least) lagged values of y, p, and m.

From (2) and the serial independence of ε_t it follows that

$$Ep_t|\Omega_{t-1} = Em_t|\Omega_{t-1} - Ey_t|\Omega_{t-1} ,$$

an expression which makes clear the dependence of the parameters characterizing expectations formation on the parameters of the money feedback rule.

From (1), we have

$$Ey_t|\Omega_{t-1} = \lambda y_{t-1} .$$

Substituting the preceding two equations and equation (2) into (1) gives

$$y_t = \gamma(m_t - y_t - \varepsilon_t - Em_t|\Omega_{t-1} + Ey_t|\Omega_{t-1}) + \lambda y_{t-1} + u_t$$

$$(1+\gamma)y_t = \gamma(m_t - Em_t|\Omega_{t-1}) - \gamma\varepsilon_t + (\gamma+1)\lambda y_{t-1} + u_t$$

or

$$(20) \qquad y_t = \frac{\gamma}{1+\gamma}(m_t - Em_t|\Omega_{t-1}) - (\frac{\gamma}{1+\gamma})\varepsilon_t + \frac{1}{1+\gamma}u_t + \lambda y_{t-1}.$$

Equation (20) is an analogue for this model of the reduced form equation (6) in our ad hoc expectations model. The difference is that the parameters of the money control rule explicitly appear in (20) by virtue of the presence of the term $Em_t|\Omega_{t-1}$. Thus, although versions of (20) will "resemble" (6), those versions are now predicted to have parameters that depend on the choice of monetary rule.

To find the optimal rule under (20), we can try to continue to follow our old advice: set $Ey_t|\Omega_{t-1} = y^*$. But from (20), we have that

$$Ey_t|\Omega_{t-1} = \lambda y_{t-1} ,$$

<u>regardless</u> of the parameters of the money supply rule. So in this model, the bias squared is <u>independent</u> of the parameters of the money supply rule. From (20), it follows that the variance of y_t around its conditional mean of λy_{t-1} is minimized by setting

$$m_t = Em_t|\Omega_{t-1} ,$$

as before. (Policy rules should be deterministic and involve no surprises, a result which emerges in both this section and the previous one.) We have therefore established the following stochastic neutrality theorem that characterizes our model: one deterministic feedback rule on the basis of the information set Ω_{t-1}, which is common to the public and to the authority, is as good as any other deterministic feedback rule. That is, the mean squared error is simply not a function of the parameters determining the systematic (forecastable) part of the money supply. Via deterministic feedback rules, the monetary authority is powerless to combat the business cycle (the serial correlation in y_t). This result is the antithesis of our earlier results rationalizing activist Keynesian policy rules.

The reader is invited to verify that the truth of the neutrality theorem is not dependent on the particular information set assumed. It will continue to hold for any specification of Ω_{t-1} so long as the public and the authority share the same information set.

The preceding results provide a (weak) defense for following rules without feedback. Simple x-percent growth rules do as well as any deterministic feedback rules, and dominate rules with a stochastic component.

Two features of the model formed by (1), (2), and (3') account for the neutrality result. The first is the assumption that the public's expectations are "rational" and that the public and the policy authority share the same information set. The second feature is that the system embodies the natural rate hypothesis, which is to say, supply decisions are homogeneous of degree zero in prices and expected prices. Abandoning either of these hypotheses will cause the conclusions of the neutrality theorem to fail.

Which View Does the Evidence Favor?

We have set out two logically consistent arguments, one in favor of rules with feedback, one arguing that rules with feedback are no better than rules without feedback and may be worse if they introduce into policy what agents perceive as noise from the point of view of their information sets. What evidence permits one to choose between these two views? Various kinds of naive arguments that economists often bring to bear will not resolve the matter. A good example is the argument often heard, that from empirical observation we "know" tax cuts are a good countercyclical device because, say, a sustained boom followed the tax cut of 1964. The reason this argument doesn't settle anything is that souped-up versions of the model formed by (1), (2), and (3') can explain such correlations, say, between tax cuts and subsequent booms in output.[4/] Even further, it can be shown quite generally that both "neutral" and "nonneutral" forms of econometric models are compatible with literally any observed patterns of the correlation in the data.[5/]

From the point of view of evaluating the case for rules with feedback, it is essential to verify the invariance assumption that is so critical to the argument. To have confidence in the argument, there should be evidence that reduced forms of estimated macroeconometric of the form (6) have remained invariant across changes in policy regimes in the past. Such tests have been performed too rarely in the past, but usually with results pointing to failure of the invariance assumption for various key reduced form equations.[6/] More tests of this kind are needed to settle the issue. In the other direction, it can be shown that the neutrality argument in effect assumes invariance across regimes of the parameters of the reduced form as written in a different form than (6)--in the case of our little model, equation (20). To assign any relevance to the neutrality theorem, evidence for the invariance of the reduced form as written in this way should be available.

As of this date, the evidence on this issue is very fragmentary and somewhat mixed.

Exercises

1. Consider the system formed by (1), (2), and (3') where (1) is replaced by

$$y_t = \gamma(p_t - {}_tp_{t-1}^*) + \lambda y_{t-1}$$

$$+ \alpha({}_tp_{t-1}^* - p_{t-1}) + u_t$$

where $\alpha > 0$ and where u_t and ε_t have the same properties assumed in the text. Calculate the optimal feedback rule for m_t, using the objective function in the text. Show that the neutrality theorem fails.

2. Consider the system formed by (1), (2), and (3') with (3') replaced by

$${}_tp_{t-1}^* = E[p_t|\Omega_{t-1}] + \xi_t$$

where ξ_t is a mother-in-law term, reflecting random deviations from rationality and satisfying $E[\xi_t|\Omega_{t-1}] = 0$. Assume that the public and the government share the same information set. Find the optimal control rule for m_t. Does the neutrality theorem hold?

Should the Monetary Authority Use Interest or Money as Its Instrument?

Consider the following macroeconomic model:

(1) $y_t = \gamma(p_t - {}_t p_{t-1}^*) + \lambda y_{t-1} + u_t$ (Phillips curve) $\gamma > 0$

(2) $m_t - p_t = y_t + br_t + \varepsilon_{1t}$ portfolio balance curve $b < 0$

(3) $y_t = c(r_t - ({}_{t+1}p_{t-1}^* - p_t)) + \varepsilon_{2t}$ "IS curve," $c < 0$.

Here r_t is the interest rate. Here u_t, ε_{1t}, and ε_{2t} are each serially independent stationary random processes with means of zero. Thus, $E\, u_t | \Omega_{t-1} = E\, \varepsilon_{1t} | \Omega_{t-1} = E\, \varepsilon_{2t} | \Omega_{t-1} = 0$. We assume that u, ε_1, and ε_2 have finite variances.

The monetary authority has the option of using a feedback rule on previous information for setting r_t, and letting m_t be whatever it must be to achieve portfolio balance at that r_t; or alternatively, of setting m_t via a feedback rule on previous information, letting r_t be whatever it must to equilibrate the system. Which of these two alternatives should the authority choose? There has been a tendency for monetarists to advocate choosing a feedback rule on m, while some (though not all) Keynesians advocate using a feedback rule on r. Previously in our study of nonstochastic models we have seen that in some classical models if the authority pegged the interest rate the price level became indeterminate (Wicksell's observation); and that in Keynesian models and other classical models with suitable definitions of disposable income, the authority could peg the interest rate but the choice between interest and money had no consequence for the value of real GNP. That is, in the nonstochastic Keynesian model, a given level of real GNP could be achieved as well by having the authority choose a suitable money supply as by having it choose a suitable interest rate.

In the context of the preceding stochastic model, however, the authority does have a choice between using r or m as its instrument, a choice with substantive implications for the probability distribution of y. For the case of _ad hoc_ expectations, we leave the details to be worked out in the following exercise:

Exercise: Complete the model by assuming the _ad hoc_ expectations schemes

$${}_{t}p_{t-1}^{*} = v\, p_{t-1}$$

$${}_{t+1}p_{t-1}^{*} = w\, p_{t-1},$$

where v and w are parameters that remain fixed in the face of variations in the rule. Assume that the authority is interested in minimizing $E_{t-1}(y_t - y^*)^2$ by choosing a feedback rule _either_ of the form

$$r_t = H\Omega_{t-1}$$

_or_

$$m_t = G\Omega_{t-1}$$

where H and G are vectors conformable to the authority's information set Ω_{t-1}, which consists of observations on all _lagged_ endogenous and exogenous variables (and therefore on lagged values of the disturbances, too, since the authority knows the model). You can assume that u_t, ε_{1t} and ε_{2t} are pairwise uncorrelated. Then prove that:

a. Whether the authority should use the interest rate rule or the money supply rule depends on the variances of u, ε_1, and ε_2 and the slopes of the IS and LM curves.

b. The more stable is the IS curve relative to the LM curve and the steeper is the IS relative to the LM curve, the more likely is it that the authority will want to use the interest rate as its instrument.

(Poole and Bailey are useful references on the problem addressed in the preceding exercise.)

So in the presence of uncertainty and under fixed-weight, <u>ad hoc</u> expectations, the choice between use of r and m as instruments has content. Notice that if the variances of the random variables u, ε_1, and ε_2 are set to zero, the choice no longer makes a difference, which agrees with our earlier remarks about the irrelevance of the choice between using r and m as instruments in a nonrandom Keynesian model.

Now let us turn to the choice of instruments under rational expectations. We supplement (1), (2), and (3) with

(4) $${}_{t}p_{t-1}^{*} = E[p_t|\Omega_{t-1}]$$

(5) $${}_{t+1}p_{t-1}^{*} = E[p_{t+1}|\Omega_{t-1}] ,$$

where Ω_{t-1} includes the same variables in the information set of the authority, namely all lagged endogenous and exogenous variables. Using (4) and taking conditional expectations in (1) we have

(6) $$Ey_t|\Omega_{t-1} = \lambda y_{t-1}.$$

Taking conditional expectations in (2) and (3) gives

(7) $$Em_t|\Omega_{t-1} - Ep_t|\Omega_{t-1} = Ey_t|\Omega_{t-1} + bEr_t|\Omega_{t-1}$$

(8) $$Ey_t|\Omega_{t-1} = c(Er_t|\Omega_{t-1} - E(p_{t+1}-p_t)|\Omega_{t-1}) .$$

Under a money supply rule, $Em_t|\Omega_{t-1}$ is given. Of course, from (6) we know that one deterministic money supply rule is as good as any other from the point of view of the mean squared error of output. But the money rule will influence the distribution of prices. To see how the distribution of prices is determined, solve (2) and (3) for $Ep_t|\Omega_{t-1}$ to get

$$(9) \qquad Ep_t|\Omega_{t-1} = \frac{1}{1-b} Em_t|\Omega_{t-1} - \frac{1+\frac{b}{c}}{1-b} Ey_t|\Omega_{t-1} - \frac{b}{1-b} Ep_{t+1}|\Omega_{t-1} .$$

Since $b < 0$, we have that $0 < \frac{-b}{1-b} < 1$.

We can solve this difference equation in $Ep_t|\Omega_{t-1}$ in the forward direction to get

$$(10) \qquad Ep_t|\Omega_{t-1} = \frac{1}{1-b} \sum_{j=0}^{\infty} \left(\frac{-b}{1-b} \right)^j Em_{t+j}|\Omega_{t-1} - \frac{(1+\frac{b}{c})}{1-b} \sum_{j=0}^{\infty} \left(\frac{-b}{1-b} \right)^j Ey_{t+j}|\Omega_{t-1} .$$

(We could substitute $\lambda^{j+1} y_{t-1}$ for $Ey_{t+j}|\Omega_{t-1}$.)

Here we are imposing the terminal condition

$$\lim_{n\to\infty} \left(\frac{-b}{1-b} \right)^n Ep_{t+n}|\Omega_{t-1} = 0 ,$$

which has the effect of asserting that in the absence of money supply changes, agents will not expect accelerating inflation or deflation. Provided that the rule is such that

$$\sum_{j=0}^{\infty} \left(\frac{-b}{1-b} \right)^j Em_{t+j} | \Omega_{t-1}$$

converges, equation (10) determines a finite expected price level $Ep_t | \Omega_{t-1}$. Presumably, then, the money supply rule can be tailored to set $Ep_t | \Omega_{t-1}$ at some desired level.

In sum, there exists a money supply rule that delivers a finite (conditionally expected) price level. While the money supply rule is powerless for affecting the probability distribution of real output (which is again a consequence of our having assumed the natural rate hypothesis together with rational expectations), the money rule does influence the distribution of prices.

Now consider an interest rate rule, which determines $Er_t | \Omega_{t-1}$. Since $Ey_t | \Omega_{t-1} = \lambda y_{t-1}$ from (1), we have that (2) determines $Em_t | \Omega_{t-1} - Ep_t | \Omega_{t-1}$, since both $Er_t | \Omega_{t-1}$ and $Ey_t | \Omega_{t-1}$ are determined. Then (3) must determine $Ep_t | \Omega_{t-1}$. Write the conditional expectation of (3) as

$$(11) \qquad Ep_t | \Omega_{t-1} = Ep_{t+1} | \Omega_{t-1} + \frac{1}{c} Ey_t | \Omega_{t-1} - Er_t | \Omega_{t-1} .$$

The solution of this difference equation is

$$Ep_t | \Omega_{t-1} = \frac{1}{c} \sum_{j=0}^{n} Ey_{t+j} | \Omega_{t-1} - \sum_{j=0}^{n} Er_{t+j} | \Omega_{t-1}$$

$$+ Ep_{t+n+1} | \Omega_{t-1} .$$

To solve this equation for $Ep_t | \Omega_{t-1}$ requires a terminal condition in the form of an exogenously given expected price level $Ep_{t+n+1} | \Omega_{t-1}$. An increase in the value that is assigned to this terminal condition results in a one-for-one increase in $Ep_t | \Omega_{t-1}$. Thus, the expected price $Ep_t | \Omega_{t-1}$ is underdetermined by the model itself, being dependent on our having to supply a very strong terminal condition that in effect determines the price level. That is, the model itself as characterized by equation (11) is incapable of restricting the price level. Another way to put this is by observing that under an interest rate rule, the terminal condition that we have to impose to determine the $Ep_t|\Omega_{t-1}$ is very much stricter than what we had to impose under the money supply rule.

The economics behind the underdetermined expected price level is this: Under the interest rate rule, the public correctly expects that the authority will accommodate whatever quantity of money is demanded at the pegged interest rate. The public therefore expects that, ceteris paribus, any increase in p_t will be met by an increase in m_t. But that means that one $Ep_t|\Omega_{t-1}$ is as good as any other one from the point of view of being rational. There is nothing to anchor the expected price level. And this is not simply a matter of choosing the "wrong" level or rule for the interest rate. There is no interest rate rule that is associated with a determinate price level.

The preceding indeterminacy is the counterpart in this model of the nonstochastic, statics result that in a full employment model with wages and prices that are instantaneously flexible, it can happen (under suitable restrictions on the IS curve) that the price level is indeterminate if the monetary authority pegs the interest rate.

Footnotes

1/ Where the disturbances follow a normal distribution or where the objective function is quadratic and the model is linear, it is known that linear feedback rules of the form (7) are the optimal ones to follow.

2/ See Karaken, Muench, and Wallace for an extended statement of this argument.

3/ See Samuelson [ ].

4/ For an example of such a model, see my "Classical Macroeconometric Model for the United States."

5/ See Sargent, "Observational Equivalence of Natural and Unnatural Rate Theories of Macroeconomics."

6/ A good example of this kind of work is the paper by Muench, Rolnick, Wallace, and Weiler.

References

Bailey, Martin, National Income and the Price Level, 2nd edition, pp. 175-186.

Kareken, John H., Thomas Muench, and Neil Wallace, "Optimal Open Market Strategy: The Use of Information Variables," American Economic Review (March 1973).

Muench, T., A. Rolnick, N. Wallace, and W. Weiler, "Tests for Structural Change and Prediction Intervals for the Reduced Forms of Two Structural Models of the U.S.: the FRB-MIT and Michigan Quarterly Models," Annals of Economic and Social Measurement, March 1974.

Poole, William, "Optimal Choice of Monetary Policy Instruments in a Simple Stochastic Macro Model," Quarterly Journal of Economics (May 1970): 197-216.

Samuelson, Paul A., "Reflections on Recent Federal Reserve Policy," Journal of Money, Credit, and Banking (February 1970).

Sargent, Thomas, "A Classical Macroeconometric Model for the United States," Journal of Political Economy, April 1976.

Sargent, Thomas, "The Observational Equivalence of Natural and Unnatural Theories of Macroeconomics," Journal of Political Economy, June 1976.