A BAYESIAN APPROACH TO ADAPTIVE EXPECTATIONS

by

Rodney L. Jacobs and Robert A. Jones

Discussion Paper #93
University of California, Los Angeles

June 1977

We wish to thank Michael Parkin for comments on a much earlier version of this paper, and Jon Salmon and Richard Sousa for their able research assistance.

I. INTRODUCTION

Expectations about the future values of certain variables are the primary determinant of individual behavior in many economic models. For instance, expectations about future inflation rates appear in models of the demand for money during hyperinflation and in theories of the long-run Phillips Curve. Expectations about future income underlie the life-cycle approach to consumption demand. Expectations about future interest rates are utilized in explanations of the term structure of interest rates. And expectations about future sales and price levels influence price and quantity adjustment dynamics in models of disequilibrium. Yet considerably less justification is usually provided for the expectation formation hypotheses employed than for other aspects of these models.

It is usually assumed that agents forecast future levels of economic variables on the basis of the past observed values of the same variable. Within this class are the static, simple adaptive, extrapolative, regressive, and unrestricted distributed lag expectation formation schemes. Choosing among these alternatives has tended to be somewhat ad hoc rather than based on the statistical properties of the resulting forecast. An alternative hypothesis is that agents form their forecasts "rationally" on the basis of all economic variables known to them and with full knowledge of the process which generates future variables. This hypothesis has considerable theoretical appeal as a means of consistently closing a model with endogenous expectations, but has been criticized for ignoring the substantial costs to an individual agent of acquiring and processing large quantities of data. Naive forecasting rules based only on the past observations of the variable to be forecast may be "economically rational" (Darby 1976, Feige & Pearce 1976) if the value of improved prediction from more sophisticated rules is outweighed by the cost of collecting and processing the additional information.

The purpose of this paper is to characterize a multi-level adaptive expectations forecasting scheme as a Bayesian decision rule for an agent with particular beliefs about the stochastic process generating future variables. The process is sufficiently general that the constant coefficient adaptive expectations and extrapolative expectations schemes emerge as special cases, even if the agent's loss function is not quadratic in the forecast error. However, the assumption of Bayesian learning is sufficient to imply strong and testable restrictions on the adaptation coefficients in the adjustment scheme.

Section I of the paper characterizes the usual adaptive expectations scheme as a Bayesian learning rule to exposit the general approach. Section II then develops the two-level model in which an agent forms beliefs about both the level and rate of change of the variable to be forecast. Section III estimates and tests the specification of section II using survey data on inflation expectations in the United Kingdom from 1961 to 1973.

I. SIMPLE ADAPTIVE EXPECTATIONS

Suppose an individual desires to forecast a variable p_t which he believes has the stochastic structure[1]

$$(1) \qquad \begin{aligned} p_t &= \bar{p}_t + u_t \\ \bar{p}_t &= \bar{p}_{t-1} + v_t \end{aligned}$$

where u_t and v_t are independent white noise processes with zero means and variances μ and ν respectively. For example, if p_t is an observed market price in period t then we could interpret $\bar{p}_t$ as the true market price, u_t as an observation error, and v_t as the drift in the price level or inflation since the previous period. Alternatively we could interpret u_t as a transistory shock and v_t as a permanent shock to the price level.[2] We assume that the individual's beliefs about $\bar{p}_t$, before observing p_t, can be represented by a prior probability distribution which is normally distributed with mean p_t^e and variance s_t. This distribution incorporates information conveyed by past observations of p, knowledge of the structure of the stochastic process given by (1) and any other factors previously affecting expectations about $\bar{p}_t$. The forecast problem is to characterize an individual's probabilistic beliefs about the price p_{t+1} to be observed in the <u>next</u> period, given his prior probabilistic beliefs about $\bar{p}_t$ and the observed value of p_t.

Upon observing the current price p_t, the individual will revise his beliefs about the current true price $\bar{p}_t$. A standard application of Bayes' Theorem (DeGroot, p. 167) reveals that the conditional probability distribution of $\bar{p}_t$ given p_t is normal with mean $p_t^{e'}$ and variance s_t' where

$$p_t^{e'} = p_t^e + \lambda_t(p_t - p_t^e),$$

(2)

$$1/s_t' = 1/s_t + 1/\mu,$$

and

$$\lambda_t = \frac{s_t}{\mu + s_t}.$$

The mean of the individual's subjective distribution on $\bar{p}_t$ is adjusted by some fraction λ_t of the difference between the observed p_t and the mean of his prior expectations about p_t, while the variance of his subjective distribution is decreased. The application of Bayes' rule leads to an adaptive expectations adjustment rule; however, in the general case given by equations (2) the value of s_t, and consequently λ_t, will vary with time. The constant coefficient adaptive expectations rule is, therefore, not necessarily a Bayesian forecasting rule. Turnovsky (1969) applies Bayes' rule to a model where the true price level remains fixed ($\nu = 0$) and obtains an adaptive expectations forecasting rule. He indicates that the constant coefficient rule follows only if the variance of the observation error or transistory shock μ falls geometrically with time. This could be the case if p_t was the average of a sample of prices taken during period t and if the size of that sample was increasing geometrically with time. We wish to derive the constant coefficient rule for the more plausible situation where $\bar{p}$ varies with time but μ need not decline.

With $\nu > 0$ what will be the individual's probabilistic beliefs about next period's true market price $\bar{p}_{t+1}$ after observing p_t? Since $\bar{p}$ will change by v_{t+1}, which has zero mean and variance ν, his prior on $\bar{p}_{t+1}$ will be normal with mean p_{t+1}^e and variance s_{t+1} where

$$(3)\qquad \begin{aligned} p^e_{t+1} &= p^{e'}_t = p^e_t + \lambda_t(p_t - p^e_t) \\ s_{t+1} &= s'_t + \nu = \frac{\mu s_t}{\mu + s_t} + \nu . \end{aligned}$$

The variance equations of (2) and (3) can be solved to obtain the stable non-linear difference equation

$$(4)\qquad s_{t+1} - s_t = \frac{- s_t^2}{\mu + s_t} + \nu$$

which describes how s_t varies with time. The equilibrium value of s_t can be computed by setting $s_{t+1} = s_t$ in equation (4) to obtain

$$(5)\qquad s = \frac{\nu}{2} + \sqrt{\left(\frac{\nu}{2}\right)^2 + \nu\mu} \; .$$

As s converges to the value of equation (5), the adaptation coefficient asymptotically approaches the constant

$$(6)\qquad \lambda = \frac{s}{\mu + s} = \frac{-\nu}{2\mu} + \sqrt{\left(\frac{\nu}{2\mu}\right)^2 + \frac{\nu}{\mu}} \; .$$

Equation (6) is identical to the expression obtained by Muth who sought that linear function of (an assumed infinite number of) past observations which provides the minimum mean square forecast error predictor of next period's price. The constant coefficient adaptive adjustment rule is therefore the asymptotic form of a Bayesian revision rule.

The value of λ in equation (6) depends only on the ratio of the two variances involved and varies from 0 to 1 as ν/μ varies from 0 to ∞. As μ increases, p_t becomes a more unreliable indicator of the current true price $\bar{p}_t$. The smaller value implied by equation (6) then puts less weight on p_t. As ν increases, past information becomes less relevant because the current true price may have drifted a considerable distance from previous true price levels. In this case, the larger value of λ implied by equation (6) puts more weight on the current observation. Although the mean of the individual's prior beliefs about the price to be observed next depends only on the ratio ν/μ and past observations, the same is not true of his subjective variance. The (asymptotic) variance of his prior distribution on p_t (including observation error) is $s + \mu$, and thus increases with both ν and μ.

The fact that the individual's prior distribution on next period's price approaches a normal distribution with <u>constant</u> variance and a mean which is given by a constant coefficient adaptive rule permits us to say something about the individual's Bayes decisions over time. Suppose the individual must choose some action at time t which we can associate with a real number x_{t+1} and whose payoff depends on p_{t+1}. The choice is made so as to minimize the expected value of some loss function $L(p_{t+1}, x_{t+1})$. Further assume that the loss function (omitting time subscripts for the moment) depends on x only through its difference with p. That is

$$(7) \qquad L(p, x) = G(p-x) + H(p)$$

An example of this sort of loss function is the usual squared forecast error where x is the agent's forecast of next period's price. A more interesting example is to interpret x as the reservation trade price of a futures contract for delivery of a commodity whose future spot price is p. The speculator's profit is proportional to p-x, and G(p-x) represents his utility associated with the corresponding profit level. It can be readily verified that a translation of the individual's subjective probability distribution on p induces an equal translation of the expected loss minimizing value of x for the above loss function. In other words, $x^* = p^e + k$

where p^e is the mean of the individual's subjective probability distribution over p and where k depends on the higher moments of the distribution and particulars of the loss function.

The importance of the above point rests with the fact that individuals' probabilistic beliefs cannot be directly observed but their actions based on those beliefs can. Suppose an agent with loss function (7) believes he is operating in an environment described by the stochastic process (1). What relationship should we observe between his actions x* and actual prices p over time? If the individual has observed p for a sufficient time, then the variance of his prior for each period is close to its asymptotic value of $s + \mu$, and hence $x^*_{t+1} - p^e_{t+1} = k$ is constant over time. Substituting $x^* - k$ into (3) for p^e implies we should observe

$$x^*_{t+1} = x^*_t + \lambda(p_t - x^*_t) + \lambda k \qquad (8)$$

Although the agent's actions do not reveal unbiased estimates of his mean beliefs about the next period's price, his actions would follow a constant coefficient adaptive scheme with the addition of a constant term.

In this section we have demonstrated (at the cost of assuming normality of all random variables) that the constant coefficient adaptive adjustment rule is asymptotically an undominated Bayesian revision rule for forming beliefs about the future values of variables generated by the stochastic process (1) rather than just the minimum-mean-square-forecast-error predictor from the more limited class of linear constant-coefficient functions of past observations. Moreover this adaptive behaviour carries over to the optimal Bayes decisions of the individual for a larger class of loss functions than the quadratic, permitting the hypothesis to be tested even when it is suspected that, because of risk-aversion or other factors, an agent's behaviour does not provide an unbiased indicator of his mean probabilistic beliefs.

II. MULTI-LEVEL ADAPTIVE EXPECTATIONS

Many economic variables relevant for individual decisions appear to follow definite trends over time. In the presence of trends the simple adaptive rule of the previous section will lead individuals to consistently underpredict or overpredict the future value of p_t. Because of the consistent prediction errors, a rational individual would quickly reject the working hypothesis that (1) represents the stochastic process generating p_t. A more general process, which admits the possibility of trends which change over time, is

$$
\begin{aligned}
p_t &= \bar{p}_t + u_t \\
\bar{p}_t &= \bar{p}_{t-1} + \bar{\pi}_t + v_t \qquad (9) \\
\bar{\pi}_t &= \bar{\pi}_{t-1} + w_t
\end{aligned}
$$

where u_t, v_t, and w_t are independent white noise processes with zero means and variances μ, ν, and ω respectively. p_t is the observed price, $\bar{p}_t$ the current true market price, $\bar{\pi}$ the trend or inflation in the true price since the last period, u_t an observation error or transistory shock in $\bar{p}_t$, v_t a permanent shock in $\bar{p}_t$ or transistory shock in $\bar{\pi}_t$, and w_t a permanent shock in $\bar{\pi}$. If $\bar{\pi}_o$ and ω are equal to zero we have the stochastic process given by equations (1). If ν equals zero the process is identical to that used by Nerlove and Wage (1964) to discuss the optimality of adaptive forecasting. The asymptotic form of the individual's Bayesian revision rule will be derived in precisely the same manner as in the previous section. However, the problem is complicated by the fact that the individual holds joint beliefs about both $\bar{p}$ and $\bar{\pi}$.

Assume the individual's beliefs about the pair $(\bar{p}_t, \bar{\pi}_t)$ prior to observing p_t can be represented by a joint probability distribution which is normal with mean (p_t^e, π_t^e) and covariance matrix $\Sigma_t = [s_{ij}]_t$. It is shown in the Appendix that the individual's posterior probability distribution on $(\bar{p}_t, \bar{\pi}_t)$ after observing p_t is joint normal with mean $(p_t^{e'}, \pi_t^{e'})$ and covariance matrix Σ_t' given by

$$p_t^{e'} = p_t^e + \lambda_1(p_t - p_t^e),$$

(10)
$$\pi_t^{e'} = \pi_t^{e'} + \lambda_2(p_t - p_t^e), \quad \text{and}$$

$$\Sigma_t' = \frac{\mu}{\mu + s_{11}} \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} + \frac{|\Sigma|}{\mu} \end{bmatrix}.$$

The adaptation coefficients are (omitting time subscripts on elements of Σ)

(11)
$$\lambda_1 = \frac{s_{11}}{\mu + s_{11}} \quad \text{and}$$

$$\lambda_2 = \frac{s_{12}}{\mu + s_{11}}.$$

The elements of Σ and consequently λ_1 and λ_2 generally change with time; however, as sampling continues they converge to constant values.

The individual's prior distribution on $(\bar{p}_{t+1}, \bar{\pi}_{t+1})$ is obtained from his posterior distribution on $(\bar{p}_t, \bar{\pi}_t)$ given above and from his beliefs about the process generating next period's prices given by (9). Specifically, his prior on next period's mean price level and inflation rate will be normal with mean

(12)
$$p_{t+1}^e = p_t^{e'} + \pi_t^{e'} = p_t^e + \lambda_1(p_t - p_t^e) + \pi_{t+1}^e$$

$$\pi_{t+1}^e = \pi_t^{e'} = \pi_t^e + \lambda_2(p_t - p_t^e)$$

and covariance matrix

(13)
$$\Sigma_{t+1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \Sigma_t' \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} + \nu \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \omega \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

The variation of Σ with time can be determined by substituting the expression for Σ_t' given by equations (10) into equation (13). This substitution yields three (since Σ is symmetric) independent difference equations which describe the convergence of Σ to its asymptotic value. Setting $\Sigma_t = \Sigma_{t+1} = \Sigma$ gives three independent second degree equations for the asymptotic values of s_{11}, s_{12}, and s_{22}. It is shown in the Appendix that these equations reduce to

$$s_{12} = \omega^{\frac{1}{2}}(\mu + s_{11})^{\frac{1}{2}}$$

(14) $$s_{22} = \omega^{\frac{1}{2}} s_{11}/(\mu + s_{11})^{\frac{1}{2}}$$

$$(\mu + s_{11})^2 - \omega^{\frac{1}{2}}(\mu + s_{11})^{\frac{3}{2}} - (2\mu + \nu)(\mu + s_{11}) - \mu\omega^{\frac{1}{2}}(\mu + s_{11})^{\frac{1}{2}} + \mu^2 = 0.$$

The last equation could be solved numerically for s_{11} if the parameters μ, ν, ω were known, and the remaining elements s_{12} and s_{22} could be determined from s_{11}.

Can anything be said about the values of λ_1 and λ_2 without knowing the values of μ, ν, and ω? Surprisingly, the values of these adaptation parameters are confined to a remarkably small region just by the structure of the model. The non-negativity of s_{11} follows from the positive-definiteness of Σ, while that of s_{12} follows from (14). That λ_1 must lie between 0 and 1, and that λ_2 must be non-negative then follows from (11). Furthermore, if we utilize $\lambda_1 = s_{11}/(\mu + s_{11})$ and $\lambda_2 = s_{12}/(\mu + s_{11}) = \omega^{\frac{1}{2}}/(\mu + s_{11})^{\frac{1}{2}}$ from the first equation of (14), the last equation of (14) can be written in terms of λ_1, λ_2 and solved for λ_2 to yield

(15) $$\lambda_2 = \frac{\lambda_1^2 - (1 - \lambda_1)\nu/\mu}{2 - \lambda_1} \leq \lambda_1$$

The area below the line $\lambda_2 = \lambda_1^2/(2 - \lambda_1)$ in Figure 1 depicts the region to which λ_1 and λ_2 are restricted by the structure of the model. The values of μ, ν, and ω which correspond to being on various boundaries are also indicated. Many of the naive expectation formation schemes hypothesized by various writers can be viewed as special cases of our model. With $\mu = \omega = 0$ $(\lambda_1 = 1, \lambda_2 = 0)$ and $\bar{\pi}_0 = 0$ we have static expectations where the individual believes that whatever price prevailed this period will prevail next period. With $\omega = 0$ $(\lambda_1 \leq 1, \lambda_2 = 0)$ we have adaptive expectations on p. With $\nu = 0$ $(\lambda_2 = \lambda_1^2/(2 - \lambda_1))$ we have the Theil-Nerlove-Wage adaptive scheme. With $\mu = 0$ $(\lambda_1 = 1, \lambda_2 \leq 1)$ the individual has simple adaptive expectations on π, where $\pi_t = p_t - p_{t-1}$ is the observed drift in p_t. With $\omega = \nu = 0$ $(\lambda_1 = \lambda_2 = 0)$ the individual follows a regressive expectation scheme in which he expects next period's price to fully revert to some fixed trend line with slope $\bar{\pi}_0$. Finally, with $\mu = \nu = 0$ $(\lambda_1 = \lambda_2 = 1)$, the adaptation scheme reduces to the extrapolative expectations hypothesis $p_{t+1}^e = p_t + (p_t - p_{t-1})$.

As we previously noted, the simple adaptive scheme of Section I will yield consistently biased forecasts when there is a trend in p_t. The model of equations (12) solves this problem by having individuals update a forecast of both the price level and its drift. Consider the simple case where ω and π are initially zero and π then jumps to a higher constant value. The individual would begin to forecast a value of p_{t+1}^e which was consistently less than p_{t+1}. As a result, he would continue to change his estimate of the drift until π^e converged to the new value of $\bar{\pi}$ and the forecasts no longer had any consistent errors. In general, π^e would overshoot the true trend and would converge to $\bar{\pi}$ through a damped oscillation or be critically damped to $\bar{\pi}$. The path of π^e would depend on the stability properties of the model. It is easy to show that the expectations scheme is stable over the entire region of Figure 1 and will converge through a damped oscillation if the values of λ_1 and λ_2 lie above the dashed line defined by $\lambda_2 = 2 - \lambda_1 - 2\sqrt{1 - \lambda_1}$.

The procedure used to obtain a specification of expectation formation in this section can be extended to higher order stochastic processes generating p_t. For example, if the trend drifts over time at a rate $\bar{d}_t$ an individual may reject (9) and choose to form beliefs as if the process generating prices is

$$
\begin{aligned}
p_t &= \bar{p}_t + u_t \\
\bar{p}_t &= \bar{p}_{t-1} + \bar{\pi}_t + v_t \\
\bar{\pi}_t &= \bar{\pi}_{t-1} + \bar{d}_t + w_t \\
\bar{d}_t &= \bar{d}_{t-1} + z_t
\end{aligned}
\tag{16}
$$

where u_t, v_t, w_t, and z_t are independent white noise processes. The asymptotic rule by which the individual updates his joint prior on $\bar{p}$, $\bar{\pi}$, and $\bar{d}$ is

$$
\begin{aligned}
p^e_{t+1} &= p^e_t + \lambda_1(p_t - p^e_t) + \pi^e_{t+1} \\
\pi^e_{t+1} &= \pi^e_t + \lambda_2(p_t - p^e_t) + d^e_{t+1} \\
d^e_{t+1} &= d^e_t + \lambda_3(p_t - p^e_t)
\end{aligned}
\tag{17}
$$

in which $1 \geq \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$. This particular specification is among those estimated in the following section. We will be particularly interested in the case where λ_1 equals unity. In this instance the model becomes

$$
\begin{aligned}
\pi^e_{t+1} &= \pi^e_t + \lambda_2(\pi_t - \pi^e_t) + d^e_{t+1} \\
d^e_{t+1} &= d^e_t + \lambda_3(\pi_t - \pi^e_t)
\end{aligned}
\tag{18}
$$

and is equivalent to updating π^e and d^e from observations on π.

III. EMPIRICAL RESULTS

Testing a model of expectations is difficult because variables generated by the model are inherently not observable. The standard approach is to imbed the expectations equations in a model which attempts to explain other phenomena such as the demand for money during hyperinflation. Testing the model then involves a joint test, and it is impossible to untangle specification error in the expectations equations from specification error in the remainder of the model. An alternative approach is provided by the various survey data on expectations such as the Carlson and Parkin data for the United Kingdom.[3] Although we will use the Carlson/Parkin data to test the model developed above, it should be realized that we have not solved the problem of joint hypothesis testing but merely changed its form. We now jointly test the model of expectations and the method employed to compute quantitative expectations data from the qualitative survey results.

For the period 1961 to 1973 a monthly Gallup Poll survey was conducted in the United Kingdom to ascertain individual beliefs about the movement of prices over the next six months. A sample of approximately 1,000 individuals were asked to indicate whether they expected prices to go up, go down, stay the same, or didn't know. Each month at least 50% of the individuals expected prices to rise and in most months some individuals expected prices to fall. Carlson and Parkin assume that individuals have a probability distribution over expected future prices. If the mean of this distribution is some threshold amount greater than the current price the individual responds that prices will go up. If the mean of this distribution is some threshold less than the current price the individual responds that prices will go down. The fraction of responses falling in each of these categories is then used to generate an index of expectations. The index is arbitrarily scaled to equate the average expectations index and the average actual rate of inflation for the thirteen year period. We assume that this monthly index represents data $\hat{\pi}^e$ on the average expected rate of inflation over the next six months. Because of the manner in which the index is scaled, we assume that the observed prices are the logarithm of the monthly price index. We then regard π^e_{t+1} as the theoretical expected rate of inflation over the next month.

Testing the expectation model involves solving equation system (17) for the theoretical forecasts as a function of time. To obtain an analytic solution we define the column vectors

$$X_t = \begin{bmatrix} p^e \\ \pi^e \\ d^e \end{bmatrix}_{t+1} \quad \text{and} \quad Z_t = \begin{bmatrix} \lambda_1+\lambda_2+\lambda_3 \\ \lambda_2+\lambda_3 \\ \lambda_3 \end{bmatrix} P_t$$

and write the model in matrix form

(19) $$X_t = AX_{t-1} + Z_t,$$

where

$$A = \begin{bmatrix} 1-\lambda_1-\lambda_2-\lambda_3 & 1.0 & 1.0 \\ -\lambda_2-\lambda_3 & 1.0 & 1.0 \\ -\lambda_3 & 0.0 & 1.0 \end{bmatrix}$$

The solution of this system of difference equations is given by

(20) $$X_t = A^t X_o + \sum_{\tau=0}^{t} A^{(t-\tau)} Z_\tau,$$

where X_o is the vector of initial conditions and the unit of time is a month.

The matrix A^t can be computed from

$$A^t = M \begin{bmatrix} \rho_1^t & 0 & 0 \\ 0 & \rho_2^t & 0 \\ 0 & 0 & \rho_3^t \end{bmatrix} M^{-1}$$

where ρ_i, the eigenvalues of A, and M, the matrix of associated eigenvectors, depend on the λ_i. If the eigenvalues lie within the unit circle, the solution is stable and consists of a series of damped sinusoidal terms. The analytical solution obtained from the eigenvalue representation is computationally burdensome. We adopt the simple, but equivalent, procedure of recursively generating the state variables. Starting from the initial conditions X_0 we numerically compute X_1 from equation (19). Given X_1 we can then compute X_2 in the same manner. By repeated use of equation (19), the time history of X_t can be generated at successive points in time.

Given the model solution, the initial conditions vector X_0 and the model parameters λ_1, λ_2 and λ_3 can be adjusted to find the set for which the theoretical forecasts best fit the data. A problem arises because the data $\hat{\pi}_t^e$ are not equivalent to any of the theoretical forecast variables. The data are expectations of the average rate of inflation over the next six months, while π_{t+1}^e is the theoretical forecast over the next month. Because of the drift d_{t+1}^e, forecasts further into the future will differ from π_{t+1}^e. If π_{t+6}^e denotes the average theoretical forecast of the rate of inflation over the next six months, then $\pi_{t+6}^e = \pi_{t+1}^e + 2.5\ d_{t+1}^e$. Initial estimation results indicated that this additional complication was unnecessary. We obtained approximately the same estimates whether the forecast was π^e_{t+1} or π^e_{t+6}. As a result, we have employed the computationally simpler approach of using π^e_{t+1} as the theoretical expected rate of inflation. To determine the initial conditions and model parameters we could then minimize the sum square of residuals $\Delta\hat{\pi}_t = \hat{\pi}^e_t - \pi^e_{t+1}$.[4] This loss function involves some problems with convergence because it is possible to generate reasonable solutions for π^e_{t+1} which involve radical departures of the predicted price series p^e_t from the actual price series. For instance, the actual price may be considerably larger than the predicted price for most of the data interval. The larger differences between predicted and actual prices would be compensated for by adjustments in the parameters λ_1, λ_2 and λ_3. To eliminate these pathological cases from consideration we need to impose the restriction that p^e_t not deviate radically from p_t. The simple procedure we follow is to give some weight to the residuals $\Delta p_t = p_t - p^e_t$. We, therefore, select the initial conditions and model parameters which minimize SSR $= \Sigma\Delta\hat{\pi}^2_t + \delta\Sigma\Delta p^2_t$ where δ is a scale factor. The scale factor was arbitrarily chosen so that $\Sigma\Delta\hat{\pi}^2_t$ contributed over 90% of the total SSR. The estimation results are not particularly sensitive to variations in the scale factor. The model is nonlinear in the initial conditions and parameters so the minimization of SSR requires an iterative algorithm. We have used the technique of Marquardt (1963) which involves an interpolation between the Taylor series and gradient methods.

Before discussing the model estimation we want to examine Figure 2 which depicts the expected rate of inflation data and the actual rate of inflation. The monthly rate of inflation is very noisy making it difficult to establish trends so we have plotted a yearly moving average.[5] It is clear that the model will be unable to reproduce certain aspects of the expectations data. There is a tremendous jump in expectations in November of 1967. This increase in expectations is unrelated to changes in the actual rate of inflation and persists for approximately six months before returning to a level consistent with the actual rate of inflation. The jump in expectations was apparently caused by the announced devaluation of the pound. After this episode the expectations data becomes much noisier although the actual inflation rates do not exhibit this pattern. From late 1971 on, the expected rate of inflation remains consistently below the actual rate of inflation. For this period, the average value of $\hat{\pi}^e$ is approximately 2% below the average value of π. This difference probably results from the way the survey data were transformed into a quantitative index of the expected rate of inflation.

Because of the erratic behavior of the expectations data after October of 1967, the model was initially estimated from the data prior to the devaluation. These estimation results are presented in the first row of Table 1. The results suggest that λ_1, is not significantly different from unity, and this hypothesis is verified by the estimation results presented in the second row of Table 1. We would, thus, have obtained equivalent results if we assumed that the rate of inflation was observed and that individuals update an expected rate of inflation and its drift. The model differs significantly from the standard adaptive expectations model which ignores the drift in the rate of inflation. Estimation results for the standard model, which consists of the constraints $\lambda_1 = 1.0$ and $\lambda_3 = 0.0$, are presented in the third row of Table 1. The results indicate that λ_3 is significantly different from zero.[6]

Our next step is to fit the entire data set to test the explanatory power of the model after the devaluation. We assign zero weight to the six values of $\hat{\pi}^e$ after October of 1967 because the model is clearly incapable of reproducing the sudden jump in expectations. Weighting these observations zero prevents their spurious residuals from dominating the estimation process. An equivalent procedure would have been to use dummy variables for these months. We use a single dummy variable for the period August of 1971 to December of 1973. The dummy variable π^* represents a constant decrease in the expected rate of inflation generated by the model and thus corrects for the consistently low values of $\hat{\pi}^e$ during this period. An equivalent approach would have been to compute a new data series $\hat{\pi}^e_t + \pi^*$.

The first row of Table 2 presents the estimation results. The parameters for the entire data interval do not appear significantly different from the parameters obtained from the initial data interval. This hypothesis is tested in the second row of Table 2 where λ_1, λ_2 and λ_3 and the initial conditions are constrained to the values obtained from the estimation prior to the devaluation. The complete estimation does not provide significantly better results than are obtained by extrapolating the initial estimation results.[7] Figure 3 illustrates the ability of π^e_{t+1} to reproduce the data series $\hat{\pi}^e_t$. For π^e_{t+1} we have plotted the value generated by the model and have, therefore, increased the data after July of 1971 by the constant π^*. We see that the theory is able to reproduce the general trend of the $\hat{\pi}^e_t$ series but not its fluctuations.[8] This is because the large swings in $\hat{\pi}^e_t$ are unrelated to movements in the monthly rate of inflation. It is interesting to note that the model provides a better forecast of the future rate of inflation than does $\hat{\pi}_t^e$. The Gallup Poll asked about expected price movements over the next six months so we have compared $\hat{\pi}_t^e$ and π_{t+1}^e with the average rate of inflation over the next six months. The survey data gives an average (in absolute value) forecast error of 3.4% while the average error of the model is only 2.8%.[9]

Our analysis of the expectations data differs markedly from that of Carlson and Parkin. They estimate the simple adaptive expectations model

$$\pi_{t+1}^e = \pi_t^e + \beta_1(\pi_t - \pi_t^e), \tag{21}$$

as well as the second order error learning model

$$\pi_{t+1}^e = \pi_t^e + \beta_1(\pi_t - \pi_t^e) + \beta_2(\pi_{t-1} - \pi_{t-1}^e), \tag{22}$$

and conclude that expectations formation is a different process in the low inflation years prior to June of 1967 compared with the remaining high inflation years. Our model is able to provide a unified explanation of these two periods because of the drift d_t^e in the expected rate of inflation.[10] The drift term captures aspects of the data which are not reproduced by the ad-hoc second order model of equation (22). Carlson and Parkin also conclude that, except for the month of the devaluation, there appear to be no other factors affecting the formation of expectations. This is contrary to our findings of a lingering influence of the devaluation which is unrelated to the actual π_t, the increased noise in the data after the devaluation, and the need for a dummy variable to correct the expectations data after September of 1971. To explain these differing views we need to examine the solution of equation (21). The analytic solution is

$$\pi_{t+1}^e = \pi_0^e(1-\beta)^{t+1} + \beta\sum_{\tau=0}^{t}(1-\beta)^{t-\tau}\pi_\tau \tag{23}$$

where π_0^e is the forecast before observing π_0. Given the model solution we could iteratively determine the values of π_1^e and β which minimize $\Sigma(\hat{\pi}_t^e - \pi_{t+1}^e)^2$. The results of this estimation for the entire data set are given in the first row of Table 3. The model fits the $\hat{\pi}^e$ data poorly and has significantly larger residuals than the model of Table 2 which includes the drift in the rate of inflation and the dummy variable after July of 1971.

Carlson and Parkin do not estimate the model solution represented by equation (23). Instead they substitute past data values for past theoretical values in equation (21) to obtain

$$(24) \qquad \pi_{t+1}^e = \hat{\pi}_{t-1}^e + \beta(\pi_t - \hat{\pi}_{t-1}^e)$$

This equation is now linear in β and can be estimated using standard regression techniques. The estimation is greatly simplified; however, this simplification is obtained by the questionable procedure of ignoring the difference between past theoretical values π_t^e and past data $\hat{\pi}_{t-1}^e$. Because of the substantial errors in variables problems associated with the data, the procedure will bias the estimation results. The procedure also tends to insure that a dynamic model accurately fits the data even if its structure is incorrectly specified. To illustrate this last point, consider the estimate of equation (24) presented in row two of Table 3. In direct contrast to our previous results, these estimation results seem to indicate that the simple adaptive expectations model accurately describes the dynamics of expectations formation. Figure 4 further illustrates differences between the two results. The model solution of equation (24) appears to accurately reproduce the data while the opposite is true for the solution of equation (23). These differing results are due to the mixing of data and theoretical values in equation (24). The substitution of $\hat{\pi}^e$ for past theoretical values prevents the model solution from deviating significantly from the data and thus insures a high value R^2. This is clearly illustrated by the data after the devaluation. The model is incapable of reproducing the large jump in expectations and this is clearly indicated by the solution of equation (23); yet, equation (24) has no problem after the first devaluation point. With equation (24) there is a large forecast error for November of 1967. This forecast error is eliminated by December because the substitution of past $\hat{\pi}^e$ for past π^e automatically draws the model back to the data series. The same process occurs after September of 1971 where the expectations data is consistently below the actual rate of inflation. This period creates no problem for equation (24) because the substitution of $\hat{\pi}^e_{t-1}$ for π^e_t keeps drawing the model back to the data.

IV. CONCLUSIONS

Compared with the simple adaptive model, we find that the ability of past price changes to explain inflation expectations is significantly enhanced if we assume individuals form beliefs about trends in the inflation rate. Furthermore the model permits a unified view of the expectations adjustment process during periods both of relatively constant and of rising inflation; there is no evidence of a change in the relationship between expectations and realized inflation. There are still large unexplained fluctuations, to be sure. But it is not clear that these result from individuals being more rational than we assume: The expectations generated by the model provide better predictions of subsequent actual inflation than does the survey data itself. Thus the model appears to have some empirical usefulness.

It may be argued that the model leading to the specification estimated merely pushes the level of ad hocery back one step -- from arbitrary choice of a naive forecasting rule to arbitrary choice of a naive stochastic view of the economy. But the model is one of the simplest capable of rationalizing the common human tendency to look for and extrapolate trends, and it does provide a basis for choice among competing specifications. The credibility of a particular naive expectations rule in a given application might be more easily assessed by examining the circumstances in which it would be rational than by examining properties of the rule itself.

REFERENCES

Anderson, T. W. An Introduction to Multivariate Statistical Analysis. New York: Wiley, 1958.

Carlson, J. A., and Parkin, J. M. "Inflation Expectations." Economica (May 1975): 123-138.

Darby, M. R. "Rational Expectations Under Conditions of Costly Information." Journal of Finance 31 (June 1976): 889-895.

DeGroot, M. H. Optimal Statistical Decisions. New York: McGraw-Hill, 1970.

Feige, E. L., and Pearce, D. K. "Economically Rational Expectations." Journal of Political Economy 84 (June 1976): 499-522.

Fleming, J. S. Inflation. London: Oxford University Press, 1976.

Marquardt, D. W. "An Algorithm for Least-Squares Estimation of Non-linear Parameters." J. SIAM 11 (June 1963): 431-441.

Muth, J. F. "Optimal Properties of Exponentially Weighted Forecasts." Journal of the American Statistical Association 55 (June 1960): 299-306.

__________. "Rational Expectations and the Theory of Price Movements." Econometrica 29 (July 1961): 315-335.

Nerlove, M., and Wage, S. "On the Optimality of Adaptive Forecasting." Management Science 10 (January 1964): 207-224.

Severn, A. K. "Further Evidence on the Formation of Price Expectations." Quarterly Review of Economics and Business (Winter 1973): 27-35.

Theil, H., and Wage, S. "Some Observations on Adaptive Forecasting." Management Science 10 (January 1964): 198-206.

Turnovsky, S. J. "A Bayesian Approach to the Theory of Expectations." Journal of Economic Theory 1 (August 1969): 220-227.

_______________. "Empirical Evidence on the Formation of Price Expectations." Journal of the American Statistical Association 65 (December 1970): 1441-1454.

APPENDIX

A. Derivation of Equation (10):

Since the individual's prior on $(\bar{p}_t, \bar{\pi}_t)$ is normal with mean (p_t^e, π_t^e), variance $\Sigma_t = [s_{ij}]_t$, and he believes the transitory component u_t of the next observation to be uncorrelated with any past random variables that may have influenced his current prior, it follows that his joint prior on $(\bar{p}_t, \bar{\pi}_t, p_t)$ is normal with mean and covariance matrix (omitting time subscripts):

(A.1) $$E[\bar{p}, \bar{\pi}, p] = (p^e, \pi^e, p^e)$$

$$\text{Cov}[\bar{p}, \bar{\pi}, p] = \left[\begin{array}{cc|c} \multicolumn{2}{c|}{\Sigma} & s_{11} \\ & & s_{12} \\ \hline s_{11} & s_{21} & (s_{11}+\mu) \end{array}\right] = \left[\begin{array}{cc|c} s_{11} & s_{21} & s_{11} \\ s_{12} & s_{22} & s_{12} \\ \hline s_{11} & s_{21} & \mu+s_{11} \end{array}\right]$$

A standard result on normally distributed random variables (T. W. Anderson, 1958, Theorem 2.5.1) provides that the conditional distribution of $(\bar{p}, \bar{\pi})$ given p is normal with mean

(A.2) $$\begin{pmatrix} p^{e'} \\ \pi^{e'} \end{pmatrix} = E\left[\left(\begin{matrix} \bar{p} \\ \bar{\pi} \end{matrix}\right) \Big| \; p\right] = \begin{pmatrix} p^e \\ \pi^e \end{pmatrix} + \begin{pmatrix} s_{11} \\ s_{12} \end{pmatrix} (\mu+s_{11})^{-1}(p-p^e)$$

from which is obtained

(A.3) $$p^{e'} = p^e + \left(\frac{s_{11}}{\mu+s_{11}}\right)(p-p^e)$$

$$\pi^{e'} = \pi^e + \left(\frac{s_{12}}{\mu+s_{11}}\right)(p-p^e).$$

The conditional covariance matrix of $(\bar{p}, \bar{\pi})$ given p is

$$\text{(A.4)} \quad \Sigma' = \text{Cov}(\bar{p},\bar{\pi}|p) = \Sigma - \begin{pmatrix} s_{11} \\ s_{12} \end{pmatrix} (\mu+s_{11})^{-1}(s_{11} \ s_{21})$$

$$= \begin{bmatrix} s_{11} & s_{21} \\ s_{12} & s_{22} \end{bmatrix} - (\mu+s_{11})^{-1} \begin{bmatrix} s_{11}^2 & s_{11}s_{21} \\ s_{11}s_{12} & s_{12}s_{21} \end{bmatrix}$$

$$= \frac{\mu}{\mu+s_{11}} \begin{bmatrix} s_{11} & s_{21} \\ s_{12} & s_{22} + \frac{|\Sigma|}{\mu} \end{bmatrix}$$

in which $|\Sigma| = s_{11}s_{22} - s_{12}s_{21}$.

B. Derivation of Equation (13):

Equation (12) requires that in a steady state

$$\text{(B.1)} \quad \Sigma = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \Sigma' \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} + \nu \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \omega \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Substituting the expression in (A.4) for Σ' into (B.1) and multiplying out the matrix expressions then yields the three independent equations (since Σ is symmetric)

$$\text{(B.2)} \quad \text{(a)} \quad s_{11}(\mu+s_{11}) = \mu(s_{11}+2s_{12}+s_{22}) + |\Sigma| + \omega(\mu+s_{11}) + \nu(\mu+s_{11})$$

$$\text{(b)} \quad s_{12}(\mu+s_{11}) = \mu(s_{12}+s_{22}) + |\Sigma| + \omega(\mu+s_{11})$$

$$\text{(c)} \quad s_{22}(\mu+s_{11}) = \mu s_{22} + |\Sigma| + \omega(\mu+s_{11})$$

Using $|\Sigma| = s_{11}s_{22} - s_{12}s_{21}$ and the fact that $s_{12} = s_{21}$ permits (c) to be solved for

$$\text{(B.3)} \quad s_{12} = \omega^{1/2}(\mu+s_{11})^{1/2}.$$

Subtracting (c) from (b), solving for s_{22}, then making use of (B.2) provides

(B.4) $$s_{22} = \frac{s_{11}s_{12}}{\mu+s_{11}} = \frac{\omega^{\frac{1}{2}}s_{11}}{(\mu+s_{11})^{\frac{1}{2}}} .$$

Finally, substituting these expressions for s_{12} and s_{22} into (a) and collecting like powers of $(\mu+s_{11})$ gives

(B.5) $$s_{11}^2 = (2\mu+s_{11})\omega^{\frac{1}{2}}(\mu+s_{11})^{\frac{1}{2}} + \nu(\mu+s_{11})$$

which can be alternatively written as a fourth degree equation in $(\mu+s_{11})^{\frac{1}{2}}$

(B.6) $$(\mu+s_{11})^2 - \omega^{\frac{1}{2}}(\mu+s_{11})^{\frac{3}{2}} - (2\mu+\nu)(\mu+s_{11}) - \mu\omega^{\frac{1}{2}}(\mu+s_{11})^{\frac{1}{2}} + \mu^2 = 0$$

Notice that this becomes a quadratic in $(\mu+s_{11})$ if $\omega = 0$ as in section I.

FOOTNOTES

[1] This is the same structure as used by Muth (1960) to discuss the optimal properties of exponentially weighted forecasts.

[2] p may be any variable whose stochastic structure is described by (1), including functions of the variable to be forecast such as the logarithm of a price level, rate of change of a price level, et cetera.

[3] The Livingston survey data for the U.S. has also been used by a number of investigators such as Turnovsky and Severn. We are using the Carlson/Parkin data because the monthly series allows a careful study of the dynamic specification of the expectations equations.

[4] The model of equations (16) is non-stochastic and the estimation strategy is equivalent to assuming that the error structure is dominated by errors in the observations $\hat{\pi}_t^e$. This estimation method was adopted because of significant problems with the data set which are discussed later in this section.

[5] The price series used to generate the model solution and the yearly moving average is the monthly index of United Kingdom retail prices published in the Employment and Productivity Gazette.

[6] We assume that the test statistic $v = \left(\frac{T-K}{q}\right)\left(\frac{SSR^+ - SSR}{SSR}\right)$, where SSR^+ is the constrained SSR, T is the number of observations, K is the number of fitting parameters, and q is the number of constraints, is distributed as

F (q, T-K). The test for $\lambda_1 = 1.0$ gives a value of $v \simeq 0.0$ when the 95% significance level for F(1, 76) is 3.98. The test for $\lambda_1 = 1.0$ and $\lambda_3 = 0.0$ gives a value of v = 9.6 when the 95% significance level for F(2, 76) is 3.13.

[7] The six constraints give a value of v = 1.1 when the 95% significance level for F(6, 149) is 2.17.

[8] The noisiness of the $\hat{\pi}$ series may be due in part to measurement error. Carlson and Parkin construct $\hat{\pi}$ to estimate the mean over the population of the median inflation rates of the individuals' subjective distributions. An assumption regarding the form of this distribution over the population was necessary, and they chose normality for computational reasons. However misspecification of the distribution would introduce measurement error into $\hat{\pi}$ as an indicator of the population average expected inflation rate. The fact that in several periods no survey respondents expected prices to fall is unlikely given the long tail of the normal distribution, and supports the possibility of this type of misspecification.

[9] If we were to select model parameters which give the minimum squared forecast error, the prediction errors of the model would be considerably reduced.

[10] Fleming (1976, 58-68) proposes a "change of gear" model of expectations adjustment which might perform similarly. He suggests that individuals switch back and forth between simple adaptive forecasting of p, $\pi = \Delta p$

and $d = \Delta^2 p$, choosing the lowest order difference of the p series which does not currently have a time trend.

Table 1

Expectations Model Fit to Monthly Data Prior to Devaluation

(1/61-10/67)

Type of Fit	λ_1	λ_2	λ_3	R^2	SSR
No Constraints	0.957	0.0133	0.0024	0.298	59.09
Constraint $\lambda_1 = 1.0$	1.0	0.0137	0.0025	0.298	59.09
Constraints $\lambda_1=1.0$ and $\lambda_3=0.0$	1.0	0.0290	0.0	0.103	74.02

Notes: The value of R^2 is for $\hat{\pi}^e$ alone while SSR is the weighted combination. The sum square of residuals for $\Delta\hat{\pi}$ are 54.31, 54.31 and 69.37 respectively. The initial conditions estimated for the unconstrained case are $p_0^e = 4.56$ (the actual $\ln(p_t) = 4.56$), $\pi_0^e = 0.0024$ (2.88%/year compared to $\hat{\pi}_0^e = 2.26\%$/year), and $d_0^e = -0.00003$.

Table 2

Expectations Model Fit to Data Set 1/61-10/67 and 5/68-12/73

Type of Fit	λ_1	λ_2	λ_3	π^* %/year	R^2	SSR
No Constraints	0.845	0.0307	0.0016	2.39	0.562	274.3
Parameters from Row 1 of Table 1	0.957	0.0133	0.0024	2.86	0.542	286.5

Notes: The value of R^2 is for the entire $\hat{\pi}^e$ series including the 6 points excluded from the fit. SSR is for the weighted combination of $\Delta\hat{\pi}$ and Δp residuals. The values of SSR for the $\Delta\hat{\pi}$ residuals alone are 265.60 and 277.61 respectively. The values of SSR for the $\Delta\hat{\pi}$ residuals including the six omitted points are 466.47 and 485.43 respectively.

Table 3

Simple Adaptive Expectations Model Fit to Data Set 1/61 to 12/73

Type of Fit	β	R^2	SSR
Equation (23)	0.059	0.205	599.7
Equation (24)	0.038	0.685	237.2

Figure 1

Relationships Between λ_1 and λ_2

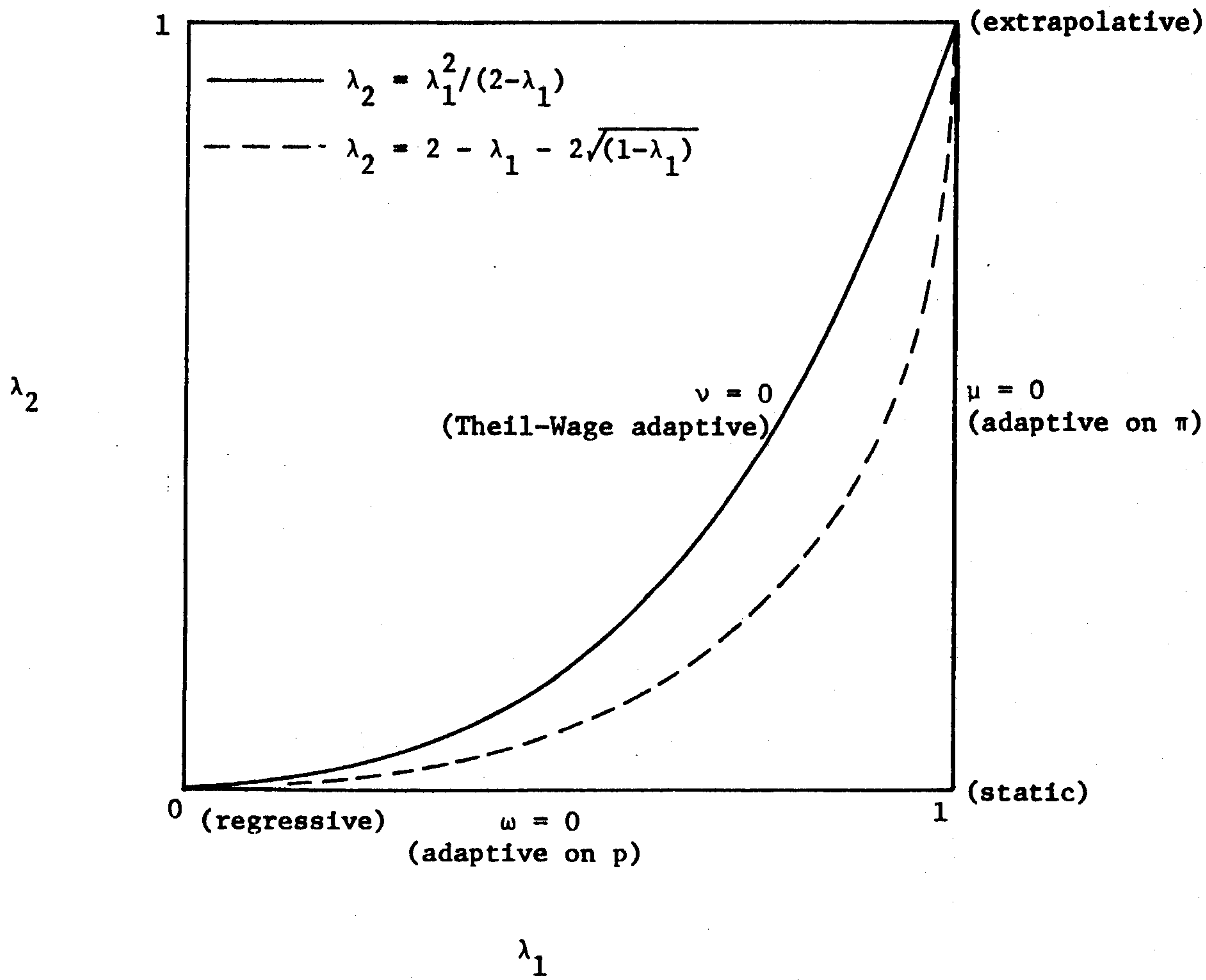

Figure 2

$\hat{\pi}_e$ and π

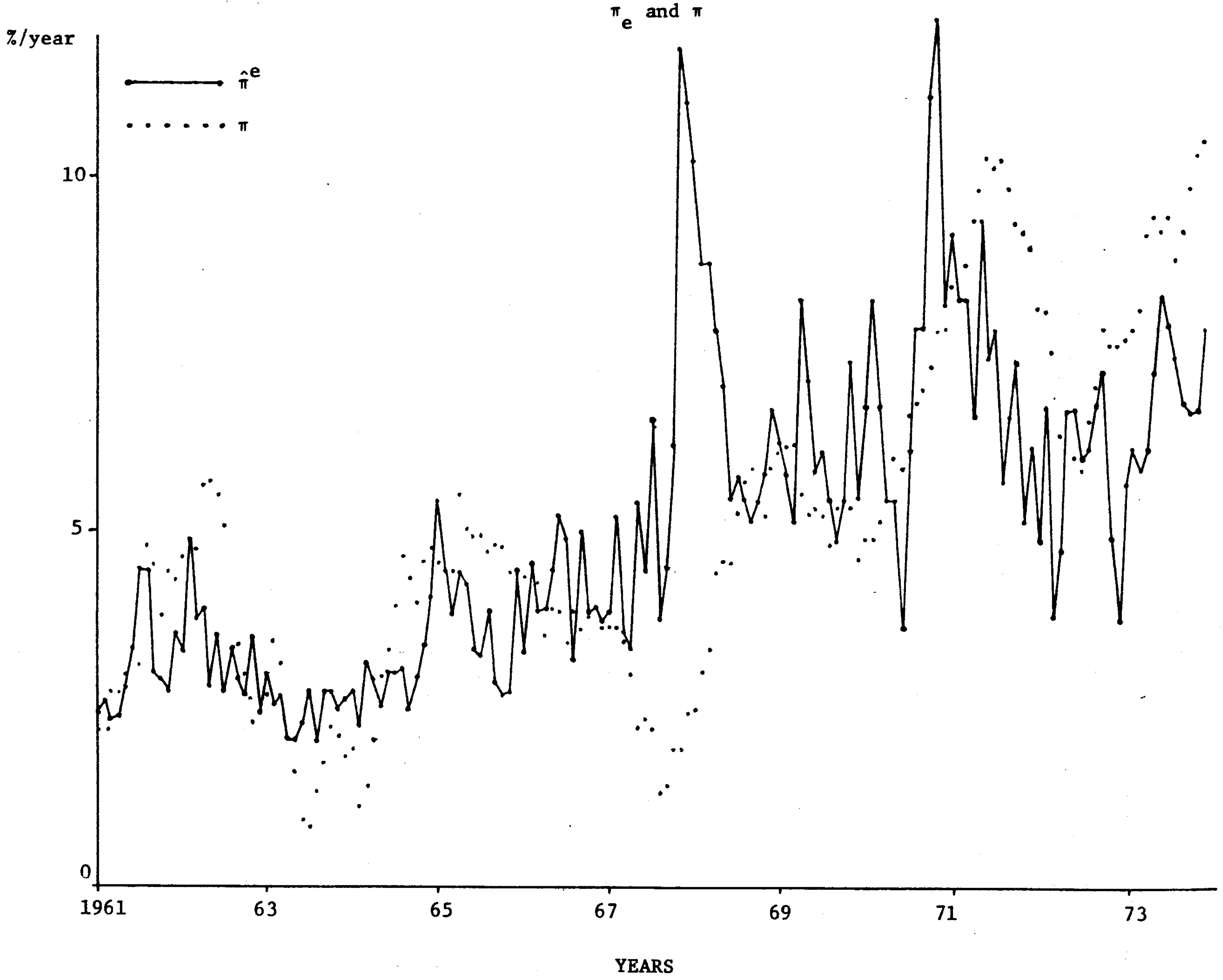

Figure 3

$\hat{\pi}_e$ and π^e_{t+1}

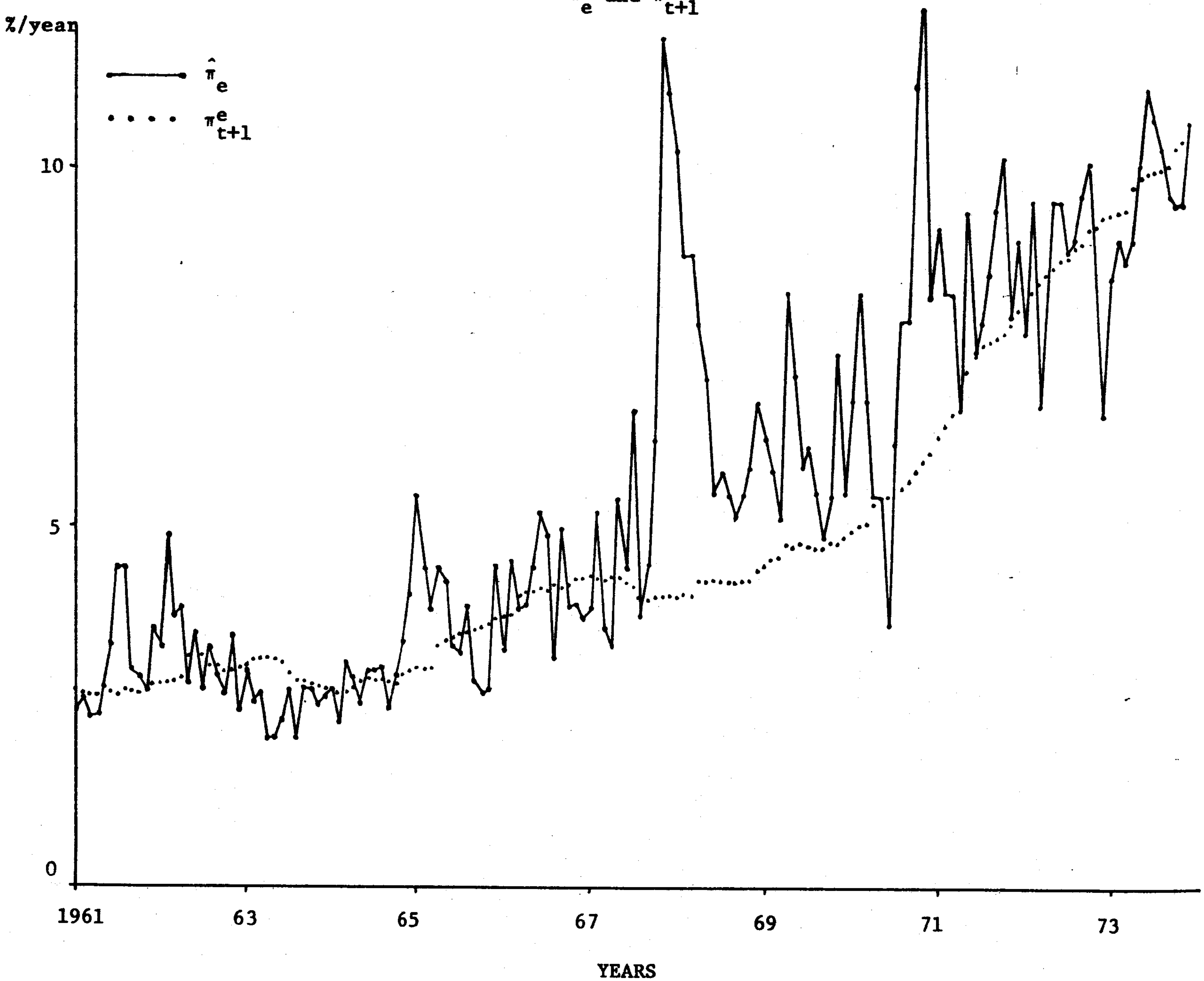

Figure 4

$\hat{\pi}^e$ and π^e_{t+1}