Inertia of Forward-Looking Expectations*

Stephen Morris
Princeton University
smorris@princeton.edu

Hyun Song Shin
Princeton University
hsshin@princeton.edu

January 2006

Abstract

Forward-looking expectations can exhibit considerable inertia when agents are differentially informed about the future path of fundamentals. Iterated average expectations of events in the far future behave like adaptive expections that put weight on outcomes in the distant past.

*Paper prepared for the session on "Information Diffusion in Macroeconomics" at the 2006 Boston meetings of the American Economic Association.

Macroeconomic variables exhibit inertia. On the face of it, such inertia sits uncomfortably with the behavior of rational, forward-looking agents who form expectations on the basis of the best information available at the time. Christopher Sims (2003, sect. 8) offers a three-fold taxonomy of attempts to rationalize inertia. The first is the idea of Robert Lucas (1973) and Edmund Phelps (1970) that agents face a signal extraction problem, so that their actions react only partially to shifts in fundamentals due to the residual uncertainty. The second is the sticky expectations approach of N. Gregory Mankiw and Ricardo Reis (2002), where otherwise rational and forward-looking agents receive information with some delay.[1] The third is the rational inattention approach favored by Sims himself, where agents with information processing constraints choose the optimal coarsening of their information (Sims (2003)).

Our approach is a variation on the signal extraction theme, but has implictions for the sticky expectations and rational inattention approaches, also. As explained recently by Michael Woodford (2003), the original contribution of Lucas (1973) had limited success in explaining the persistence of macroeconomic aggregates due to the feature that the underlying fundamentals are fully revealed after a delay of one period. However, without this simplifying feature, the complexity of the problem increases rapidly as expectations of others' expectations become relevant in one's calculations, (Robert Townsend (1983), Phelps (1983)).[2]

Tractability aside, the approach nevertheless holds much promise. When agents are differentially informed, events in the distant past are "more com-

[1] See also Carroll (2003) who provides empirical estimates of shifts in expectations.

[2] This approach was further developed by Thomas Sargent (1991). More recently, there has been a revival of the approach in macroeconomics. See Woodford (2003), Klaus Adam (2003), Jeffery Amato and Hyun Song Shin (2006), Christian Hellwig (2002), Guido Lorenzoni (2003), Takashi Ui (2003), and Kristoffer Nimark (2005).

mon knowledge" than events in the recent past. Not only has enough time elapsed for the agents to ascertain the facts in the distant past, but more importantly, enough time has elapsed for each agent to be more confident that *other* agents have managed to ascertain the facts, too. When common knowledge is important (as in coordination problems, for example) events in the distant past therefore take on significance. Paradoxically, the further the agents peer into the future, the further they must look back into the past in order to establish the informational platform for their actions. To an outside observer, the aggregate action betrays all the outward signs of adaptive expectations. In what follows, we illustrate this principle and explain the mechanisms at work.

I. Decision Rule

We consider the decision rule:

$$a_{it} = \gamma E_{it}(\theta_t) + \beta E_{it}(a_{t+1}) \tag{1}$$

where a_{it} is agent i's action at date t, $E_{it}(.)$ is the expectation with respect to i's information set at t, β is a parameter that lies strictly between zero and one and $\gamma > 0$. We will suppose that there is a continuum of agents indexed by the unit interval $[0,1]$ and a_t (without the subscript i) is the *average* action at date t, defined as $a_t = \int_0^1 a_{it} di$. The random variable θ_t is the underlying "fundamental" quantity in the economy that individuals care about. Taking the average across individuals, the average action at date t is

$$a_t = \gamma \bar{E}_t(\theta_t) + \beta \bar{E}_t(a_{t+1}) \tag{2}$$

where $\bar{E}_t(.)$ is average of the expectations taken across all agents, defined as $\bar{E}_t(.) = \int_0^1 E_{it}(.)\, di$. Equation (2) resembles the New Keynesian Phillips curve derived from rational expectations models with sticky prices, where a_t is inflation and θ_t is marginal cost. In this context, β is the discount factor used by price-setting firms, and γ is a parameter that depends on the stickiness of prices and other fundamentals of the economy.[3]

Another way to rationalize (2) is in terms of a decision problem with spillover effects where the objective is to choose an action today that balances two objectives. The first is to choose an action that matches the realization of current fundamentals θ_t. The second is to match the average action a_{t+1} next period. For instance, the agents may be firms deciding on investment levels where a_{it} is i's investment at date t and where the reward to the agent's investment depends on the fundamentals at the time of the decision, as well as the overall investment level at this following date, due to positive spillover effects. Since β is strictly between zero and one, we can iterate (2) forward to obtain:

$$a_t = \gamma \sum_{j=0}^{\infty} \beta^j \bar{E}_t \bar{E}_{t+1} \cdots \bar{E}_{t+j} \left(\theta_{t+j}\right) \tag{3}$$

If there were no differential information the iterated expectations would collapse to the single expectation at date t. Thus, in the absence of differential information we would have

$$a_t = \gamma \bar{E}_t \left(\sum_{j=0}^{\infty} \beta^j \theta_{t+j} \right) \tag{4}$$

However, with differential information, (3) will not in general be the same as (4).[4] The impact of differential information arises as the result of the divergence of (3) from (4). We examine an example where the divergence is quite stark.

[3]See Rotemberg (1982), Gali and Gertler (1999) or Mankiw and Reis (2002). Staggered price setting in the manner of Taylor (1980) gives analogous expressions (see Fuhrer and Moore (1995)).

[4]see Morris and Shin (2002) for a simple illustration in the static case

II. Uncertainty and Information

Time is indexed by the integers $\{\cdots, -2, -1, 0, 1, 2, \cdots\}$. The fundamentals $\{\theta_t\}$ are fixed at a known value y until date 0. At date 1, there is an innovation to the fundamentals that comes from the realization of a zero-mean random variable η. Thereafter, the fundamentals are fixed at its new value forever. Thus, the fundamentals evolve in the following simple fashion.

$$\theta_t = \begin{cases} y & \text{for } t \leq 0 \\ y + \eta & \text{for } t \geq 1 \end{cases} \tag{5}$$

Most of the agents observe the realization of η immediately, but not all of them do so. Over time, even those agents who did not immediately observe the realization of η gradually learn of its realized value. At date $t \geq 1$, proportion μ_t of the agents know the true value of θ_t, where μ_t is strictly increasing in t, and $\mu_t \to 1$ as t becomes large. Thus, eventually, everyone learns the true value of fundamentals. We envisage μ_1 being close to 1, so that the informational friction is small relative to the benchmark case of perfectly informed agents.

At date $t > 1$, the informed agents know the future values of θ perfectly. Hence, if agent i is informed, $E_{it}(\theta_{t+h}) = \theta_{t+h}$ for all $h > 0$, so that

$$E_{it} \begin{bmatrix} y \\ \theta_{t+h} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y \\ \theta_{t+h} \end{bmatrix} \tag{6}$$

If agent i is one of the uninformed agents, then $E_{it}(\theta_{t+h}) = E(y + \eta) = y$, so that

$$E_{it} \begin{bmatrix} y \\ \theta_{t+h} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y \\ \theta_{t+h} \end{bmatrix} \tag{7}$$

Since proportion μ_t are informed at date t, the average expectation is

$$\bar{E}_t \begin{bmatrix} y \\ \theta_{t+h} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 - \mu_t & \mu_t \end{bmatrix} \begin{bmatrix} y \\ \theta_{t+h} \end{bmatrix} \tag{8}$$

Hence

$$
\begin{aligned}
\bar{E}_{t-1}\bar{E}_t \begin{bmatrix} y \\ \theta_{t+h} \end{bmatrix} &= \bar{E}_{t-1} \begin{bmatrix} 1 & 0 \\ 1-\mu_t & \mu_t \end{bmatrix} \begin{bmatrix} y \\ \theta_{t+h} \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 1-\mu_t & \mu_t \end{bmatrix} \bar{E}_{t-1} \begin{bmatrix} y \\ \theta_{t+h} \end{bmatrix} \qquad (9) \\
&= \begin{bmatrix} 1 & 0 \\ 1-\mu_t & \mu_t \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1-\mu_{t-1} & \mu_{t-1} \end{bmatrix} \begin{bmatrix} y \\ \theta_{t+h} \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 1-\mu_{t-1}\mu_t & \mu_{t-1}\mu_t \end{bmatrix} \begin{bmatrix} y \\ \theta_{t+h} \end{bmatrix} \qquad (10)
\end{aligned}
$$

where (9) follows from the linearity of average expectations. Iterating (10), we have:

$$
\bar{E}_1\bar{E}_2 \cdots \bar{E}_t \begin{bmatrix} y \\ \theta_{t+h} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1-\prod_{s=1}^{t}\mu_s & \prod_{s=1}^{t}\mu_s \end{bmatrix} \begin{bmatrix} y \\ \theta_{t+h} \end{bmatrix} \qquad (11)
$$

The higher order iterated expectation $\bar{E}_1\bar{E}_2 \cdots \bar{E}_t(\theta_{t+h})$ in the limiting case as $t \to \infty$ depends on the limiting property of $\prod_{s=1}^{t}\mu_s$. For illustrative purposes, let

$$
\mu_s = \alpha^{-\frac{1}{s}} \qquad (12)
$$

where $\alpha > 1$, but where α is close to 1. The fact that α is close to 1 implies that almost all agents are informed from the outset. Even those agents that start out being uninformed learn over time. However,

$$
\ln \prod_{s=1}^{t} \mu_s = -\ln \alpha \cdot \sum_{s=1}^{t} \frac{1}{s} \qquad (13)
$$

and the sum $\sum_{s=1}^{t}\frac{1}{s}$ increases without bound as t becomes large.[5] Hence, $\prod_{s=1}^{t}\mu_s \to 0$ as $t \to \infty$. In other words, when $\{\mu_s\}$ is given by (12), we have

$$
\bar{E}_1\bar{E}_2 \cdots \bar{E}_t(\theta_{t+h}) \to y \quad \text{as } t \to \infty. \qquad (14)
$$

[5]Note, for instance, that $\sum_{s=1}^{t}\frac{1}{s} = \int_1^t f(x)\,dx$ where $f(.)$ is the step function that takes value $1/n$ in the interval $[n, n+1)$, so that $\sum_{s=1}^{t}\frac{1}{s}$ is bounded below by $\ln t$.

In the process of forming progressively higher order expectations, all the information in the population is lost. The limit is the expectation of the least informed individual. Note the contrast between the case where $\mu_s = 1$ and where $\mu_s < 1$. If $\mu_s = 1$, then $\bar{E}_1 \bar{E}_2 \cdots \bar{E}_t (\theta_{t+h}) = \bar{E}_1 (\theta_{t+h}) = \theta_{t+h}$ so that expectations adjust immediately. But when $\mu_s < 1$, the situation can be very different. As seen in (14), very high order expectations do not adjust by much. In this sense, there is a discontinuity as we move from the case where everyone is fully informed to one where *virtually* everyone is fully informed.

III. Inertia

We now examine the implications of our observations for the behavior of average actions $\{a_t\}$. It is helpful to set $\gamma = 1 - \beta$ for this purpose, and denote $\mu(t, t+j) \equiv \prod_{s=t}^{t+j} \mu_s$. Let us also set $y = 0$. Then at $t \geq 1$,

$$
\begin{aligned}
a_t &= (1-\beta) \sum_{j=0}^{\infty} \beta^j \bar{E}_t \bar{E}_{t+1} \cdots \bar{E}_{t+j} (\theta_{t+j}) \\
&= (1-\beta) \sum_{j=0}^{\infty} \beta^j \mu(t, t+j) \cdot \eta \qquad (15)
\end{aligned}
$$

In the special case where all agents are fully informed, we have $\mu_s = 1$ for all s, so that (15) is just

$$
\begin{aligned}
a_t &= (1-\beta) \sum_{j=0}^{\infty} \beta^j \eta \\
&= \eta \qquad (16)
\end{aligned}
$$

for all $t \geq 1$. Actions adjust immediately, and there is no inertia whatsoever. Contrast this with the case where μ_s is given by (12). Then, the coefficient on η in (15) is the weighted average of $\{\mu(t, t+j)\}$ where the weights sum to 1. The weights are determined by the discount factor β. Note that

$\mu(t, t+j) \to 0$ as j becomes large. As the discount factor β gets closer to 1, the weight placed on smaller values of μ increases. In the limit,

$$a_t \to 0 \quad \text{as } \beta \to 1 \tag{17}$$

so that there is total inertia. For values of β close to 1, the inertia will be very large in the sense that a_t is close to its former value y.

In the New Keynesian Phillips curve interpretation of our decision rule, the discount factor β is the firm's discount factor in calculating the present value of its stream of profit. If dates are interpreted as quarters, and firms discount at around 1% per quarter, then β is approximately 0.99. Thus, values of β close to 1 would seem plausible and the limiting case (17) may be of more than just theoretical interest.

IV. Markov Chain Interpretation

Although the example we have chosen in this illustration is deliberately stark, the arguments are more general. The calculation of higher order expectations can be given a Markov chain interpretation, as we see from equation (10). The matrix:

$$\begin{bmatrix} 1 & 0 \\ 1-\mu_{t-1}\mu_t & \mu_{t-1}\mu_t \end{bmatrix}$$

is the two step transition matrix over the set $\{y, \theta_{t+h}\}$, which is the product of the one-step transition matrices:

$$\begin{bmatrix} 1 & 0 \\ 1-\mu_t & \mu_t \end{bmatrix}, \qquad \begin{bmatrix} 1 & 0 \\ 1-\mu_{t-1} & \mu_{t-1} \end{bmatrix}$$

The Markov chain is "non-homogeneous" in the sense that the transition probabilities change over time. The weights associated with higher order expectations arise from the multi-stage transition probabilities. Note that y

is an absorbing state, since if the system starts at y, it will remain there. If all agents are fully informed, then $\mu_t = 1$, so that θ would also be an absorbing state. The Markov transition would then be the trivial one associated with the identity matrix.

However, if $\mu_t < 1$ for all t, then it is possible (as in our example) that θ is a *transient state* in the Markov chain. Over long horizons, the probability that it will return to θ becomes smaller. In the limit as the horizon becomes infinite, the probability that it will be in θ goes to zero.

The discontinuity as we go from $\mu_t = 1$ to $\mu_t < 1$ arises from the fact that θ changes from being a persistent state to a transient state. Whether it does so or not depends on the limiting property of the transition probabilities in the Markov chain. In general, any random variable that is not common knowledge has the potential to become transient in the Markov chain. See Amato and Shin (2006) for more discussion of this point, and for other examples. See Franklin Allen, Morris and Shin (2002) for an example of how the Markov chain argument can be used to show that asset prices place "too much weight" on publicly available information including past prices, relative to the statistically optimal weights in forecasting the fundamentals of an asset.

V. Concluding Remarks

Forward-looking expectations can exhibit considerable inertia when agents are differentially informed about the future path of fundamentals. Iterated average expectations of events in the far future behave like adaptive expectations that put weight on outcomes in the distant past.

We have examined an example where agents' information sets are totally ordered (some are better informed than others). However, the argument extends to more realistic settings where agents' information sets are only partially ordered, and where they each have their own "window" on the world. Each window gives a partial view of the world - of a particular sector of the economy or a particular geographical region. Each agent has a view of his or her small "sliver" of the economy that, when pieced together mosaic fasion, reveals the full picture. In Friedrich Hayek's (1945, p. 519) words, information in the economy exist "solely as the dispersed bits of incomplete and frequently contradictory knowledge which all the separate individuals possess." No agent has a privilged view of the complete picture. The Lucas-Phelps island economy metaphor captures this feature. In this respect, we take the differential information literally. It's just a feature of the world that agents have their own window on the world.

In spite of the different starting points, our approach has implications for the sticky information and rational inattention approaches, also. One of the debates surrounding macroeconomic persistence is on the incidence of "backward-looking" behavior that is necessary to explain the observed degree of persistence (see Gali and Gertler (1999)). Gali and Gertler report a high figure - around 35% or so. A key message of our analysis is that a very small incidence of uninformed agents can generate large amounts of persistence. It is this *discontinuity* that is the distinctive feature of our approach. The incidence of sticky information or rational inattention that is necessary to account for the observed degree of persistence may be quite small when embedded in a world of differential information.

References

Adam, Klaus (2003) "Optimal Monetary Policy with Imperfect Common Knowledge", forthcoming in *Journal of Monetary Economics*.

Allen, Franklin, Stephen Morris and Hyun Song Shin (2002) "Beauty Contests and Iterated Expectations in Asset Markets" forthcoming in *Review of Financial Studies*.

Amato, Jeffery and Hyun Song Shin (2006) "Imperfect Common Knowledge and the Information Value of Prices" *Economic Theory*, 27, 213-241.

Carroll, Christopher (2003) "Macroeconomic Expectations of Households and Professional Forecasters," *Quarterly Journal of Economics*, 118, 269-298.

Fuhrer, J. and G. Moore (1995) "Inflation Persistence" *Quarterly Journal of Economics*, 110, 127-160.

Gali, Jordi and Mark Gertler (1999) "Inflation Dynamics: A Structural Econometric Analysis" *Journal of Monetary Economics*, 44, 195-222.

Hayek, Friedrich (1945) "The Use of Knowledge in Society" *American Economic Review*, 35, 519-530.

Hellwig, Christian (2002). "Public Announcements, Adjustment Delays and the Business Cycle." working paper, UCLA.

Lorenzoni, Guido (2003) "Imperfect Information, Consumers' Expectations and Business Cycles" working paper, MIT.

Lucas, Robert E. (1973) "Some International Evidence on Output-Inflation Tradeoffs" *American Economic Review*, 63, 326-34.

Mankiw, N. Gregory and Ricardo Reis (2002) "Sticky Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve," *Quarterly Journal of Economics*, 117, 1295-1328.

Morris, Stephen and Hyun Song Shin (2002). "The Social Value of Public Information," *American Economic Review*, 92, 1521-1534.

Nimark, Kristoffer (2004) "Calvo Pricing and Imperfect Common Knowledge: A Forward Looking Model of Rational Inflation Inertia", working paper, European University Institute,

Phelps, Edmund (1970) "Introduction" in E. S. Phelps et al. (eds) *Microeconomic Foundations of Employment and Information Theory*, pp 1- 23, Norton, New York.

Phelps, Edmund (1983) "The Trouble with "Rational Expectations" and the Problem of Inflation Stabilization" in R. Frydman and E. S. Phelps (eds) *Individual Forecasting and Aggregate Outcomes,* Cambridge University Press, New York.

Rotemberg, Julio (1982) "Monopolistic Price Adjustment and Aggregate Output" *Review of Economic Studies*, 49, 517-531.

Sargent, Thomas J. (1991) "Equilibrium with Signal Extraction from Endogenous Variables" *Journal of Economic Dynamics and Control*, 15, 245-273.

Sims, Christopher (2003) "Implications of Rational Inattention," *Journal of Monetary Economics*, 50, 665-690.

Taylor, John B. (1980) "Staggered Price Setting in a Macro Model" *American Economic Review*, 69, 108-113.

Townsend, Robert (1983). "Forecasting the Forecasts of Others," *Journal of Political Economy* 91, 546-588.

Ui, Takashi (2003) "A Note on the Lucas Model: Iterated Expectations and the Neutrality of Money" working paper, Yokohama National University http://www2.igss.ynu.ac.jp/~oui/

Woodford, Michael (2003) "Imperfect Common Knowledge and the Effects of Monetary Policy," in *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, (eds.) P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford, Princeton University Press, Princeton.