3
The Credibility Problem

Introduction

This chapter describes the basic expectational Phillips curve model, modifications of which underlie both story lines. The following chapter sets out a repeated economy version of this model. Together, these two chapters formalize the temptation to inflate unleashed by the discovery of the Phillips curve, the value of a commitment technology for resisting that temptation, and the fragility of reputational mechanisms as substitutes for commitment. Despite aliases, rational expectations is the only equilibrium concept throughout. Alterations in the timing of actors' decisions induce different economies with distinct outcomes.

These two chapters summarize rational expectations benchmarks that form the point of departure for the adaptive expectations models that follow. The concluding section of the next chapter offers my opinion of how faithfully this benchmark theory supports the triumph of natural-rate theory as an interpretation of history.

One-period economy

A government faces a credibility problem whenever it wishes to make decisions sooner than it must. Comparing outcomes under two different timing protocols reveals a credibility problem. In one, by choosing before the private sector, the government takes into account its effects on private decisions. In the other, the government decides after the private sector. The deterioration in outcomes under the second timing protocol measures the loss from the inability to commit.

Though credibility problems are intrinsically dynamic, it is possible to describe them in the context of a one-period model under different patterns of within period timing. This way of introducing the credibility problem prepares the way for multi-period analyses. I describe a version of the one-period model of Kydland and Prescott in Stokey's (1989, 1990) terms.

A government chooses a $y \in Y$. There is a continuum of private agents, each of whom sets ξ, its expectation about y. The average setting of ξ is x. To capture that each private agent solves a forecasting problem, define the one period payoff function of a private agent as

$$u(\xi, x, y) = -.5[(y-\xi)^2 + y^2]. \tag{1}$$

Given y, each private agent maximizes its payoff (solves its forecasting problem) by setting $\xi = y$. Since all private agents face the same problem, $x = \xi$. We require that $y \in Y$, $\xi \in X \equiv Y$, $x \in Y$. In our analysis of credible government policies, we shall make Y a compact subset of the real line.

To implement Kydland and Prescott's example, let (U, y, x) be the unemployment rate, the inflation rate, and the public's expectation of the inflation rate, respectively. The government's one-period payoff is

$$-.5(U^2 + y^2). \tag{2}$$

Unemployment is determined by an expectations augmented Phillips curve[1]

$$U = U^* - \theta(y - x), \quad \theta > 0. \tag{3}$$

The equation asserts that unemployment deviates from U^*, the natural-rate, only when there is surprise inflation or deflation.

[1] In 1973, Lucas used this parametric Phillips curve to approximate the restrictions that his 1972 model placed on time series of unemployment and inflation.

Substituting (3) into (2) lets us express the government's payoff as a function $r(x, y)$ defined as

$$r(x, y) = -.5[(U^* - \theta(y - x))^2 + y^2]. \tag{4}$$

We work with the following objects.

RATIONAL EXPECTATIONS EQUILIBRIUM: A triple (U, x, y) satisfying (3) and $y = x$.

GOVERNMENT (ONE-PERIOD) BEST RESPONSE: Given the public's expectation x, a decision rule $B(x) = \arg\max_y r(x, y)$ for setting y.

NASH EQUILIBRIUM: A pair (x, y) satisfying (i) $x = y$, and (ii) $y = B(x)$.

RAMSEY PROBLEM: $\max_y r(y, y)$. The *Ramsey outcome* is the value of y that attains the maximum.

BEST RESPONSE DYNAMICS: The dynamical system $y_t = B(y_{t-1})$, $y_0 \in Y$.

A rational expectations equilibrium is a U, x, y triple that lies on the Phillips curve, and for which private agents are solving their forecasting problem (i.e., they are not fooled), given x. Substituting $x = y$ into the Phillips curve (3) shows that $U = U^*$ in any rational expectations equilibrium. This identifies U^* as the natural unemployment rate.

A Nash equilibrium builds in a best response by the government, taking the state of expectations x as given, and also a response $x = y$ by the market, i.e., rational expectations for a given y. The government's best response function is

$$y = B(x) = \frac{\theta}{\theta^2 + 1} U^* + \frac{\theta^2}{\theta^2 + 1} x. \tag{5}$$

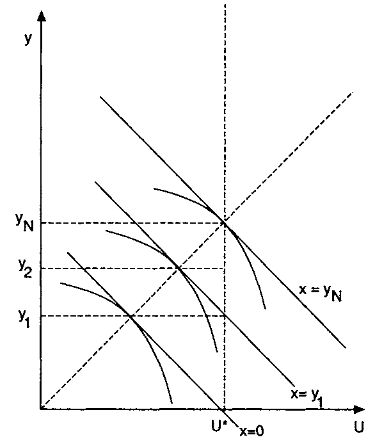

Figure 3.1. Nash equilibrium, Ramsey outcome, and best response dynamics.

The Nash equilibrium is $y^N = x^N = \theta U, U = U^*$. The Ramsey outcome is $y^R = x^R = 0, U = U^*$. Thus, $r(x^R, y^R) = -.5U^{*2}$ and $r(x^N, y^N) = -.5(1 + \theta^2)U^{*2}$.

The Nash equilibrium is supported by a timing protocol in which the government decides after the private sector sets its expectations. The Ramsey equilibrium is associated with a timing protocol in which the government chooses first, knowing that it is manipulating the private sector's expectations, because $y = x$ in a rational expectations equilibrium.

Figure 3.1 depicts the Nash equilibrium outcome and the Ramsey outcome when $\theta = 1$. Solid lines depict a family of Phillips curves for different levels of expected inflation x, with slope $\theta = -1$; curves are drawn for $x = 0, y_1, y^N$. The Nash equilibrium outcome is (U^*, y^N). The Ramsey outcome is $(U^*, 0)$. The government's best response setting for y, given x, occurs at the tangency of an indifference equation induced by (2) with the Phillips curve indexed by x.

Figure 3.2 depicts the best response function. Given that the public expects inflation x, the government's best one-period action is to set $y = B(x)$. A Nash equilibrium occurs where $y = x = B(x)$, i.e., where $B(x)$ intersects the 45 degree line. Best response dynamics convert the one-period model into a dynamic one by positing an adaptive mechanism by which x depends on the observed history of y, namely, $x_t = y_{t-1}$. This leads to the dynamics $y_t = B(y_{t-1})$ depicted in Figure 3.1. Let the system start out with $x = 0$. Then the government sets $y = y_1$. This provokes the public to set $x = y_1$, leading the government to reset $y = y_2$. The limit of this process is evidently $y = y^N$, $x = y^N$, a self-confirming situation. Thus, best response dynamics converge to a Nash equilibrium, and reinforce (U^*, y^N) as the prediction of the model without a government commitment technology.

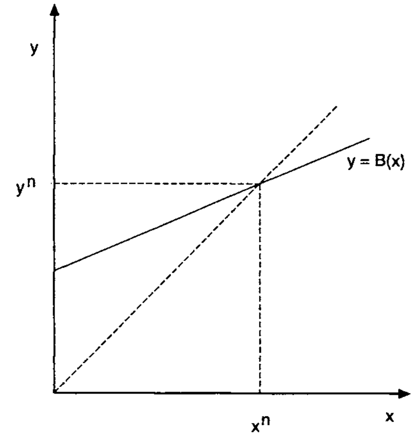

Figure 3.2. The government's best response function, $B(x)$.

A version of these best response dynamics also emerges from least squares learning.

Least squares learning converges to Nash

Least-squares learning plays a key role in this essay. The present example introduces analytical elements of least squares learning in self-referential systems, elements that reappear later in more complex settings. Least-squares learning can, like best response dynamics, converge to the Nash equilibrium outcome.

Margaret Bray (1982) studied convergence to rational expectations equilibrium of a competitive market with a one-period supply response lag in which traders formed the expected price by averaging past ones. In the same spirit, for integer $t \geq 2$, assume that x_t is the average of past y_t's, so $x_t = (t-1)^{-1} \sum_{s=1}^{t-1} y_s$. Represent x_t recursively

$$x_t = x_{t-1} + (t-1)^{-1}(y_{t-1} - x_{t-1}), \tag{6}$$

where $x_1 = 0$. Assume that actual y_t is formed from a disturbed version of the best response mapping evaluated at x_t:

$$y_t = B(x_t) + \eta_t. \tag{7}$$

Here η_t is an independent and identically distributed random term with mean zero, included to represent the government's imperfect control of inflation. Substituting (6) into (7) gives

$$x_t = x_{t-1} + (t-1)^{-1}\left[B(x_{t-1}) - x_{t-1} + \eta_t\right]. \tag{8}$$

By applying the theory of stochastic approximation, we show that the limiting behavior of x_t emerging from (8) is described by an associated differential equation[2]

$$\frac{dx}{dt} = B(x) - x. \tag{9}$$

[2] The ordinary differential equation is formed mechanically by setting $\frac{d}{dt}x = E\left[B(x) - x + \eta\right]$, where E is the mathematical expectation over the unconditional distribution of η, evaluated at a fixed x.

The rest point of this differential equation evidently satisfies $x = B(x)$, which makes it the Nash equilibrium inflation rate $x = \theta U^*$. The linearity of the best response mapping (5) makes the ordinary differential equation (the ODE) (9) linear. Define $\mathcal{M} = \frac{d}{dx}(B(x) - x)$. Evidently, $\mathcal{M} = B'(x) - 1 = -\frac{1}{\theta^2+1}$. Because $\mathcal{M} < 0$, the ODE is stable about the rest point. This brings to bear theorems described by Marcet and Sargent (1989a) that supply conditions under which the convergence of x_t to y^N occurs globally.[3] The appendix to this chapter contains an introduction to stochastic approximation.

Thus, the least squares dynamics confirm the pessimism of the best response dynamics. This pessimism undermines the triumph of natural-rate theory story, especially the triumph aspect. Given an initial condition in the form of a gold standard or Bretton Woods value of x, the best response or least squares dynamics can explain the acceleration of inflation observed in figure 1.1. But they cannot explain the stabilization that Volcker engineered.

Later I reformulate versions of least squares learning in ways designed to moderate this pessimism. But this will move us away from the triumph and toward the vindication story.

[3] See Marcet and Sargent (1995) and Evans and Honkapohja (1998b) for discussions linking $\mathcal{M}$ also to the rate of convergence of x to x^N. A necessary condition for convergence at the typical rate of $t^{.5}$ is that $\mathcal{M} < -.5$. When $\mathcal{M} \in [-.5, 0)$, x_t converges to x^N but at a slower rate governed by the absolute value of $\mathcal{M}$. See Chen (1993) and Chen and White (1993) for some results on rates of convergence for nonparametric recursive algorithms.

More foresight

Best response and least squares are out of equilibrium dynamics tacked on to a one-period economy. They force all movement through expectation formation.[4] In choosing inflation, the government forgets that the economy lasts for more than one period.

Better outcomes can occur if the government plans for the future. In subsequent chapters, I describe three different ways of modeling foresight. These impute varying amounts of rationality and predict different qualities of outcomes. The first draws from the literature on reputation or sustainable plans and attributes rational expectations to both the government and the public. This setting either confirms the pessimism emerging from Kydland and Prescott's analysis or replaces it with agnosticism: many outcomes are sustainable, ranging from repetition of the Ramsey outcome to paths much worse than repetition of the one-period Nash outcome.

A second keeps the government rational but gives the public adaptive expectations in the original Friedman–Cagan sense.[5] Depending on a comparison between a discount factor and an adaptation parameter, this setup can improve outcomes and possibly sustain repetition of the Ramsey outcome.

A third attributes adaptive behavior to both the government and the public. The outcomes depend on details of the beliefs assigned to the government. A vindication of econometric policy evaluation can emerge.

[4] The only state variable measures the public's expectations.

[5] By adopting the Cagan-Friedman specification, we will restrict the strategy space in a way to limit the set of outcomes far below those found in Chapter 4.

Appendix on stochastic approximation

An argument of Kushner and Clark (1978) shows that the ODE governs the tail behavior of the original stochastic difference equation (6). The two key components are a shift in time scale and a liberal application of averaging.

For $n \geq 0$, let $\{a_n\}_{n=0}^{\infty}$ be a positive sequence of real scalars obeying $\lim_{n\to\infty} a_n = 0, \sum_n a_n = +\infty$, $\sum_n a_n^2 < +\infty$. As an example, $a_n = \frac{1}{n+1}$ satisfies these assumptions. Rewrite (6) as

$$x_{n+1} = x_n + a_n \left[B(x_n) - x_n + \eta_n\right], \tag{10}$$

where again η_n is an independent and identically distributed process with mean zero and finite variance. To deduce the differential equation associated with the stochastic difference equation (10), we work with the transformed time scale $t_0 = 0$, $t_n = \sum_{i=0}^{n-1} a_i$. Our strategy will be to represent the solution of the difference equation (10) as a distributed lag, to transform the time scale from n to t_n, and then to interpolate the solution from the discrete points $t_n, n \geq 0$ to $t \geq 0, t \in I\!R$. In working with the extension to positive real t's, it will be convenient to use the mapping $m(t) = \max\{n : t_n \leq t\}$, which serves as an inverse mapping from t back to the original units of time n.

For $n \geq 0$, let $x^0(t_n)$ denote the solution of (10) with initial condition $x^0(0)$ at $n = 0$. Recursions on (10) yield

$$x^0(t_n) = x^0(0) + \sum_{i=0}^{n-1} a_i \left[B(x(t_i)) - x(t_i)\right] + \sum_{i=0}^{n-1} a_i \eta_i. \tag{11}$$

Following Kushner and Clark (1978), define piecewise-linear interpolated values of $x^0(t)$ for $t \in I\!R, t \geq 0$, by $x^0(t_n) = x_n$ and

$$x^0(t) = \frac{t_{n+1} - t}{a_n} x_n + \frac{t - t_n}{a_n} x_{n+1}, \quad t \in (t_n, t_{n+1}).$$

Define the piecewise-linear interpolated values $\eta^0(t)$ in an analogous fashion. The interpolated $x^0(t)$ is evidently a continuous-time stochastic process.

Kushner and Clark use the distributed lag representation of the solution (11), which is exact at the original points t_n, to motivate approximation of $x^0(t)$ by the integral equation

$$x^0(t) = x^0(0) + \int_0^t \left[B(x^0(s)) - x^0(s)\right] ds + R(t), \tag{12}$$

where $R(t)$ is an approximation error constructed by subtracting the rest of the right side of (12) from the right side of (11). The approximation error $R(t)$ evidently has two components, one associated with approximating the distributed lag in $B(x) - x$ with an integral; the other with the distributed lag in η_s in (11). Kushner and Clark want somehow to drive each of these components and therefore the approximation error R to zero.

This is not possible for small t, but it is for a large enough t. Kushner and Clark work with a sequence of left-shifted versions of (11). In particular, they define the nth left-shifted process

$$\begin{aligned} x^n(t) &= x^0(t + t_n), \quad t \geq -t_n \\ x^n(t) &= x^0(0), \quad t \leq -t_n. \end{aligned} \tag{13}$$

The left-shifted process can be represented as

$$x^n(t) = x^n(0) + \int_0^t \left[B(x^n(s)) - x^n(s)\right] ds + W_n(t) + R^n(t), \tag{14}$$

where

$$W_n(t) = \sum_{i=n}^{m(t_n+t)} a_i \eta_i, \tag{15}$$

and where $R^n(t)$ equals the difference between an interpolant of a distributed lag in $B(x) - x$ and the integral.

Kushner and Clark display technical conditions under which the two approximation errors $W_n(t)$ and $R^n(t)$ can be driven to zero as $n \to +\infty$. The key in making the random process $W_n(t)$ converge to zero is to note that it is a martingale sequence with variance $\mathrm{var}(\eta) \sum_{i=n}^{m(t_n+t)} a_i^2$ which by virtue of the assumption that $\sum_{i=0}^{\infty} a_i^2 < +\infty$ approaches zero as $n \to \infty$. The term $R^n(t)$ is sent to zero by making $a_i \to 0$ as $i \to \infty$, which drives to zero the norm of the mesh of the term $\sum_{i=0}^{N-1} a_{i+n} \left[B(x^n(t_i)) - x^n(t_i)\right]$ used to approximate the integral.[6]

By studying the limit as $n \to 0$ of the sequence of left-shifted processes in (14), Kushner and Clark establish that the limiting behavior of the original stochastic difference equation (10) is shared by the non-stochastic integral equation

$$\tilde{x}(t) = \tilde{x}(0) + \int_0^t B(\tilde{x}(s)) - \tilde{x}(s))d\,s. \tag{16}$$

Differentiating gives the ODE

$$\frac{d}{d\,t}\tilde{x}(t) = B(\tilde{x}(t)) - \tilde{x}(t). \tag{17}$$

Equation (16) or (17) is said to describe the mean dynamics of the original system (10).

Later we will study systems like (10) in which a_i does not approach zero as i grows.

[6] Note that $\{a\}_{i+n}^{N-1}$ is serving as the partition of the x-axis in the Riemann-Stieljes approximation to the integral.