Optimal Nonlinear Policy: Signal Extraction with a Non-Normal Prior

Eric T. Swanson
Board of Governors of the Federal Reserve System
eswanson@frb.gov

Abstract

The literature on optimal monetary policy typically makes three major assumptions: 1) policymakers' preferences are quadratic, 2) the economy is linear, and 3) stochastic shocks and policymakers' prior beliefs about unobserved variables are normally distributed. This paper relaxes the third assumption and explores its implications for optimal policy. The separation principle continues to hold in this framework, allowing for tractability and application to forward-looking models, but policymakers' beliefs are no longer updated in a linear fashion, allowing for plausible nonlinearities in optimal policy. We consider in particular a class of models in which policymakers' priors about the natural rate of unemployment are diffuse in a region around the mean. When this is the case, it is optimal for policy to respond cautiously to small surprises in the observed unemployment rate, but become increasingly aggressive at the margin. These features of optimal policy match statements by Federal Reserve officials and the behavior of the Fed in the 1990s.

JEL Classification: E52

Version 2.0
December 23, 2004

I thank Brian Sack, Dave Small, Volker Wieland, Larry Meyer, Kosuke Aoki, Jim Stock, Noah Williams, two anonymous referees, and conference participants at the NBER Program in Monetary Economics, the North American Meetings of the Econometric Society, the ASSA Meetings, and the European Central Bank for helpful discussions, comments, and suggestions. The views expressed in this paper, and all errors and omissions, should be regarded as those solely of the author, and are not necessarily those of the individuals listed above, the Federal Reserve System, or its Board of Governors.

1. Introduction

The literature on optimal monetary policy under uncertainty typically makes three major assumptions: 1) policymakers' preferences are well-approximated by a quadratic function, 2) the economy is well-approximated by a linear system of equations, and 3) stochastic shocks and policymakers' priors about unobserved variables are normally distributed. While these assumptions contribute much in the way of tractability and simplicity to the models and their solutions, it is important to understand the effects of relaxing these constraints on the prescriptions for optimal policy. In particular, we demonstrate that relaxing the assumption of normality in favor of a prior that is more diffuse in a region around the mean provides a very plausible source of nonlinearity in optimal policy. Relaxing priors in this way also seems to match statements by Federal Reserve officials and the Fed's behavior in the late 1990s, demonstrating that the optimal nonlinearities in this paper may be important in practice as well as in principle.

It is well known that the Linear-Quadratic-Gaussian model defined by assumptions 1–3 above yields an optimal policy response function that is linear in the observable state variables of the model. One can go beyond the linear response of policy to the state variables by relaxing any of the three assumptions. For example, Orphanides and Wieland (2000) consider a case where policymakers' preferences are non-quadratic by introducing a "zone of indifference" for inflation rates between 0 and 2 percent (motivated, for example, by some inflation-targeting central banks' official charters). Orphanides and Wieland (2000) also consider a model with a nonlinear Phillips curve with a concave-to-convex shape (i.e., a shape similar to $y = x^3$) that is like the one estimated by Filardo (1998). These approaches effectively relax assumptions 1 and 2 above, yielding an optimal nonlinear policy response to unemployment and inflation even in a world of perfect certainty, as well as a world in which disturbances and policymakers' priors are all normally distributed.

The present paper argues that relaxing the third assumption provides in many cases a more realistic model of nonlinearities in the conduct of policy. For example, a number of statements by Federal Reserve officials in the late 1990s suggest that it was primarily *uncertainty* about the economy and its natural rate of unemployment or potential out-

put that drove their behavior, rather than zone-type preferences or a significant belief in nonlinear Phillips curves. In Levy (2000), Roger Ferguson states that: "Even now as we talk about imbalances, there is an implicit short-run Phillips curve concept embedded in the discussion without necessarily saying that the unemployment rate at which inflation starts to pick up is exactly 4.5 percent, or 4.9 percent, or 5.2 percent or 5.5 percent. [I'm] in the middle in believing that there is a short-term trade-off between resource utilization and inflation, but not necessarily being wed to a specific point estimate on the short-run Phillips curve where inflation is likely to accelerate." In a similar vein, Laurence Meyer (1999) argues that: "Policymakers could attenuate the response of the real federal funds rate to declines in the unemployment rate in a region around their estimate of the NAIRU. But once the unemployment rate gets far enough below (or above) the estimated NAIRU so that confidence returns that the labor market is experiencing excess demand (or supply), then the more normal response of real interest rates to incremental declines in the unemployment rate would again become appropriate."

This paper shows that relaxing the assumption of normally-distributed priors in favor of distributions that are more diffuse in a region around the mean provides a very simple rationale for optimal nonlinear policy responses that also matches the intuition in the above statements very closely. Meyer, Swanson, and Wieland (2001) also discuss this motive for a nonlinear policy response. The present paper goes beyond that derivative work in several respects: 1) the key features of policymakers' priors that lead to an attenuated response of policy are derived and discussed; 2) this paper shows how such "diffuse-middled" priors can arise naturally in a dynamic model with structural change; 3) the persistence properties and evolution of policymakers' beliefs over time is investigated, so that the dynamic behavior of beliefs in the model is explicitly considered; and 4) the present paper shows how this analysis extends to optimal policy in a forward-looking model.

The remainder of the paper proceeds as follows. Section two sets up an illustrative model of the economy and a univariate signal extraction problem which conveys the basic intuition for the results. Section three shows that the basic results continue to hold in more realistic cases by extending the basic model along three dimensions: considering dynamic behavior of the priors as they evolve over time, multivariate rather than univariate

signal extraction, and forward-looking rather than backward-looking models. Section four compares the predictions of the model to actual Fed behavior in the 1990s, and section five concludes. A mathematical Appendix derives the technical conditions under which policymakers' priors lead to policy attenuation for small surprises in unemployment and inflation.

2. A Simple Model

The main points of this paper are independent of the exact model under consideration; thus, for the purposes of clarity and illustration, it is advantageous to begin with a model that is as simple as possible while still conveying all of the relevant intuition. The following two-equation, backward-looking model serves as a useful baseline:

$$(u_t - u^*) = \theta(u_{t-1} - u^*) + \alpha(r_{t-1} - r^*) + \varepsilon_t \tag{1a}$$

$$\pi_t = \pi_{t-1} - \beta(u_{t-1} - u^*) + \nu_t \tag{1b}$$

where u_t is the unemployment rate at time t, r_t the real interest rate, π_t the inflation rate, u^* the natural rate of unemployment, r^* the natural rate of interest, and ε_t and ν_t are stochastic shocks to the system that are orthogonal to all variables dated $t-1$ or earlier. Note that we assume for simplicity that policymakers have direct control over the real interest rate r_t and that although we have stated model (1) in terms of unemployment, one can just as easily frame it in terms of output or some broad measure of capacity utilization.

Policymakers set interest rates so as to minimize a discounted sum of squared deviations of unemployment and inflation from target values:

$$\min\ (1-\delta)\ E_t \sum_{s=t}^{\infty} \delta^{s-t} \left[(\pi_s - \pi^*)^2 + \gamma(u_s - u^*)^2\right] \tag{2}$$

where π^* denotes policymakers' objective for inflation, δ is a discount factor, γ is the relative weight on unemployment stabilization, and policymakers' objective for unemployment agrees with the natural rate u^*.

An interesting feature of the model—and one that is of primary relevance for the present paper—is that u^* is never observed. Policymakers observe current and past values

of u and π, and their own past choices for r, which they use to help infer the true value of u^* by Bayesian updating. It is assumed for simplicity that the structure of model (1) and the parameters α, β, θ, and r^* are all known with certainty.

Policymakers' problem is thus a standard linear-quadratic (LQ) programming problem with an unobserved state variable, the solution to which is well known (e.g., Bertsekas (1987)):

$$r_t - r^* = a\, E_t(u_t - u^*) + b\,(\pi_t - \pi^*) \tag{3}$$

where E_t denotes the mathematical expectation conditional on all information $\mathcal{I}_t$ available at time t:

$$\mathcal{I}_t \equiv \{\alpha, \beta, \gamma, \delta, \theta, \sigma_\varepsilon^2, \sigma_\nu^2, F_{u^*|0}(\cdot), \pi^*, r^*, \pi_t, u_t, \pi_s, r_s, u_s \mid s < t\} \tag{4}$$

$F_{u^*|0}(\cdot)$ denotes policymakers' prior distribution on u^* at time 0. The constants a and b in (3) are determined by the parameters of the model and are independent of the variances σ_ε^2 and σ_ν^2 of ε and ν (the property of certainty equivalence). Solution (3) also has the well-known property of separability of estimation and control: first, the unemployment gap $(u_t - u^*)$ is estimated on the basis of all information available at time t, and second, the interest rate r_t is set based on this estimate. It is important to note that solution (3) and the properties of certainty equivalence and separability of estimation and control are completely general properties of the LQ model and do not require normality of ε, ν, or the prior on u^*.[1]

The problem of estimating u^* is inherently one of signal extraction: policymakers never observe the true value of u^*; instead, they receive noisy observations of u^* through realizations of u_t and π_t.[2] Policymakers enter period t with prior beliefs about u^*, the expected value of which is $E_{t-1}u^*$, and based on this prior and lagged variables policymakers have prior forecasts for u_t and π_t, namely:

$$E_{t-1}u_t = (1-\theta)E_{t-1}u^* + \theta u_{t-1} + \alpha(r_{t-1} - r^*) \tag{5a}$$

$$E_{t-1}\pi_t = \pi_{t-1} - \beta(u_{t-1} - E_{t-1}u^*) \tag{5b}$$

[1] See, e.g., Bertsekas (1987), pp. 102–6, 292–3. We will return to this point for the forward-looking LQ model in section 3, below.

[2] To emphasize the signal extraction properties of equations (1), u^* could be shifted from the left-hand side to the right-hand side of equation (1a), which would put equations (1) into standard form with the observed variables on the left-hand side.

Policymakers use the ex post realizations of u_t and π_t to update their beliefs about the true value of u^* via Bayesian updating. When policymakers' priors for u^* and the shocks ε and ν are all normally distributed, then this estimation stage of policymakers' problem is linear in the observed state variables u, π, and r. In order to maintain a simple, linear form for optimal policy in terms of the observable state variables, the literature on optimal monetary policy has thus maintained the assumption of normality.

However, given the uncertainty policymakers face about the natural rate of unemployment and the possibility of structural change, it seems reasonable to think that policymakers might have beliefs about u^* that, rather than being normally distributed, are more diffuse in a region around the mean. Figure 1 below presents density functions for three distributions that might be used to model policymakers' beliefs about u^* (all the distributions have been centered around a mean of 5 for concreteness and comparability):

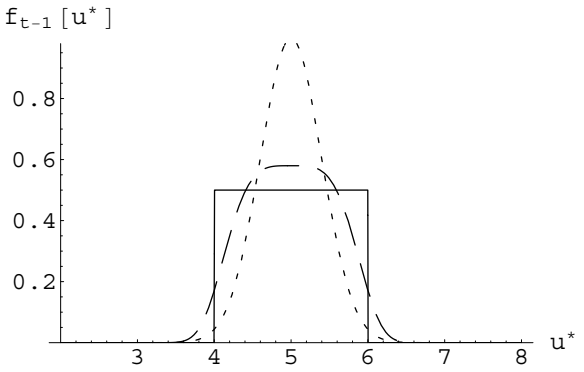


Figure 1

The short-dashed line plots a Gaussian density, which is standard in the literature. The solid line plots a uniform density over the interval $[4, 6]$, which implies a much greater degree of uncertainty about u^* in a region around the mean (although it has the unrealistic feature that policymakers are absolutely certain the true value of u^* lies neither below 4 nor above 6). The long-dashed line represents an intermediate case—it has a density that is proportional to $\exp(-0.5(u^* - 5)^4)$.[3] Throughout this paper, we will argue that these latter two "diffuse-middled" distributions are more plausible models of policymakers' beliefs than is the standard Gaussian assumption, where the term "diffuse-middled" refers

[3] The exact formula is $ke^{-\frac{1}{2}((x-5)/.8)^4}$, where the normalization constant $k = 1/((.8)\,2^{5/4}\,\Gamma(5/4))$. The (short-dashed) normal density is distributed $N(5, .16)$.

to distributions that are more diffuse in a region around the mean than is a Gaussian distribution of the same variance. In the Appendix, we provide a detailed analysis and proofs of the relationship between "diffuse-middled" priors in general and the features of optimal policy discussed below. In section 3, below, we show how such beliefs can evolve naturally in a dynamic model that allows for the possibility of structural change.

The non-normality of policymakers' priors will be reflected in their posterior estimates of u^* and the unemployment gap, $u_t - u^*$, hence in the setting of optimal policy in (3). Figure 2 illustrates this process when policymakers only have to perform univariate signal extraction, which we achieve by temporarily dropping equation (1b) from the model—while multivariate signal extraction is not unduly difficult, the figures are much simpler for the univariate case and no intuition is lost from the simplification.[4] Policymakers' estimates of u^* and $(u_t - u^*)$, which would have been linear in u_t had we assumed priors to be Gaussian, have the following functional forms when we assume the uniform prior (solid line in Figure 1):[5]

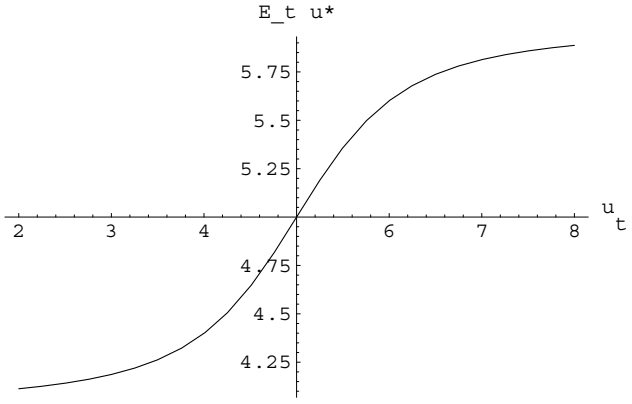


Figure 2a

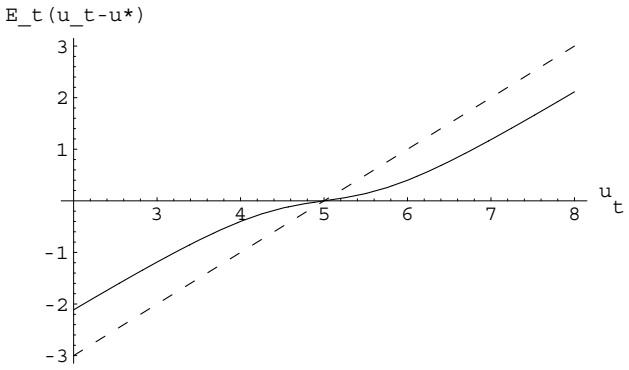


Figure 2b

Intuitively, because policymakers are so uncertain about u^* within the interval $[4, 6]$, they are very willing to revise their estimate of u^* for observations of u_t that are well inside this interval, as is evident in Figure 2a. As observed unemployment moves farther away

[4] The more general case of multivariate signal extraction is demonstrated as one of the extensions in section 3, below.

[5] In these figures, policymakers' prior forecast $E_{t-1}u_t$ equals 5 (and the expected unemployment gap $E_{t-1}(u_t - u^*)$ is zero); it is easy to show for this simple one-equation version of the model that policymakers will always have set r_{t-1} so that this is the case. The standard deviation of ε_t is set equal to 0.5, corresponding to annual U.S. data.

from policymakers' prior point estimate of 5, however, policymakers assign an increasingly smaller fraction of each increment of unemployment to u^*, this fraction approaching zero as u_t becomes more and more extreme relative to policymakers' priors. Correspondingly, the fraction of each increment of unemployment assigned to the unemployment gap, $u_t - u^*$, is close to 0 near the middle of Figure 2b, and approaches 1 as u_t moves out toward the edges of that figure.

As a result, policymakers set r_t very cautiously for small surprises in the realized unemployment rate, but respond increasingly aggressively at the margin as the surprise in unemployment becomes larger, approaching the marginal certainty-equivalent response in the limit. Thus, the theory in this paper provides a refinement of the linear signal-extraction-based motive for optimal policy attenuation discussed in Svensson and Woodford (2003) and Swanson (2004).[6] The non-normal priors in this paper accentuate the attenuation effect by increasing the variance of policymakers' priors without making the tails implausibly large; moreover, optimal policy in this paper returns to a more aggressive policy response at the margin for larger surprises, a feature of policy that seems to match the intuition offered by Fed Governor Meyer in the Introduction.

Finally, note that these basic features of the results—policy attenuation for small surprises followed by increasingly aggressive responses at the margin—are not specific to the uniform prior. In the Appendix, we prove that these properties hold for a wide class of distributions that we will refer to as "diffuse-middled"—distributions that are more diffuse in a region around the mean than a Gaussian distribution of the same variance. Swanson (2000) provides additional examples of such distributions.

3. Extensions of the Model

The simple model of the previous section illustrates the basic intuition underlying the non-normal-priors justification for optimal nonlinear policy. In this section, we extend the

[6] Sack and Wieland (2000) survey the literature on optimal policy attenuation. The motivation given by Svensson and Woodford (2003) and Swanson (2004) is based on signal extraction, as in the present paper. Note also that, unlike the present paper, Swanson (2004) takes the signal extraction problem one step further by assuming the *estimated* unemployment gap $E_t(u_t - u^*)$ is *itself* an indicator for the true underlying state variable of interest, X_t, which one might think of as "excess demand." That feature could be incorporated into the analysis of the present paper as well, but we do not do so for simplicity.

illustrative model of the previous section along three dimensions to investigate the robustness of this basic intuition to the following features: 1) dynamic evolution of policymakers' beliefs over time (including how policymakers might arrive at such "diffuse-middled" priors to begin with), 2) multivariate signal extraction, and 3) optimal monetary policy in a forward-looking model.

3.1 Dynamic Evolution of Policymakers' Beliefs

The Bayesian updating in the illustrative model above was essentially static, being computed for only a single period t given priors from period $t-1$. This section investigates two questions: First, how might policymakers have arrived at such a "diffuse-middled" prior in the first place? Second, how persistent is this diffuse-middled prior—i.e., for how many periods should the implied nonlinearities in the optimal policy be expected to last? To answer the first question, we use a simple stylized model of structural change. We then simulate the model forward to show how priors in the model evolve over time to provide an answer to the second question.

Assume that u^* is constant over time except for the possibility of structural change, which follows a Poisson process with arrival probability p each period. Conditional on a structural break arriving, a jump in u^* is drawn from a normal distribution with mean 0 and variance σ^2. Thus, model (1) is augmented with the following equation for u^*:

$$u_t^* = u_{t-1}^* + \eta_t \tag{1c}$$

$$\eta_t = \begin{cases} 0 & \text{with probability } 1-p \\ X \sim N(0, \sigma^2) & \text{with probability } \quad p \end{cases}$$

Note that the shock η has mean 0 and constant variance, so that equation (1c) preserves the underlying LQ structure of model (1).[7]

We set up initial conditions and then simulate model (1) forward over time to see how policymakers' beliefs evolve. To keep the simulation simple and focus on the dynamic evolution of beliefs, we continue to abstract away from inflation equation (1b)—which

[7] The variance of η can be allowed to vary over time, so long as this variation is exogenous (i.e., not related to any of the endogenous state variables of the model).

keeps policymakers' signal extraction problem a univariate one—and we drop interest rates from the model ($\alpha = 0$ in equation (1a)), so that policymakers just observe the economy and update their beliefs about u^* over time. These simplifications reduce the number of variables we must keep track of over time and make the simulation very clear and intuitive.

We initialize the model in period 0 with a prior distribution on u^* that is normally distributed with mean 6 and standard deviation 0.3, which is a rough calibration to the U.S. economy in the mid-1990s. We choose a normal distribution as the initial prior to demonstrate that even starting from a purely Gaussian prior, we can arrive at the more diffuse types of distributions discussed above. Moreover, after many periods without structural change, policymakers' beliefs would evolve to a distribution that is close to normally distributed anyway, so one can argue that a normal distribution is also roughly calibrated to the U.S. economy in the mid-1990s, when there had not been a substantial structural break in u^* for many years. We set $p = .01$ and $\sigma = 1$, so that a break in u^* occurs roughly once every 25 years and each break draws a jump in u^* from a $N(0, 1)$ distribution. Based on quarterly U.S. data, we set $\theta = .8$ and $\sigma_\varepsilon = .3$ in equation (1a).[8] These parameters are all known by policymakers with certainty.

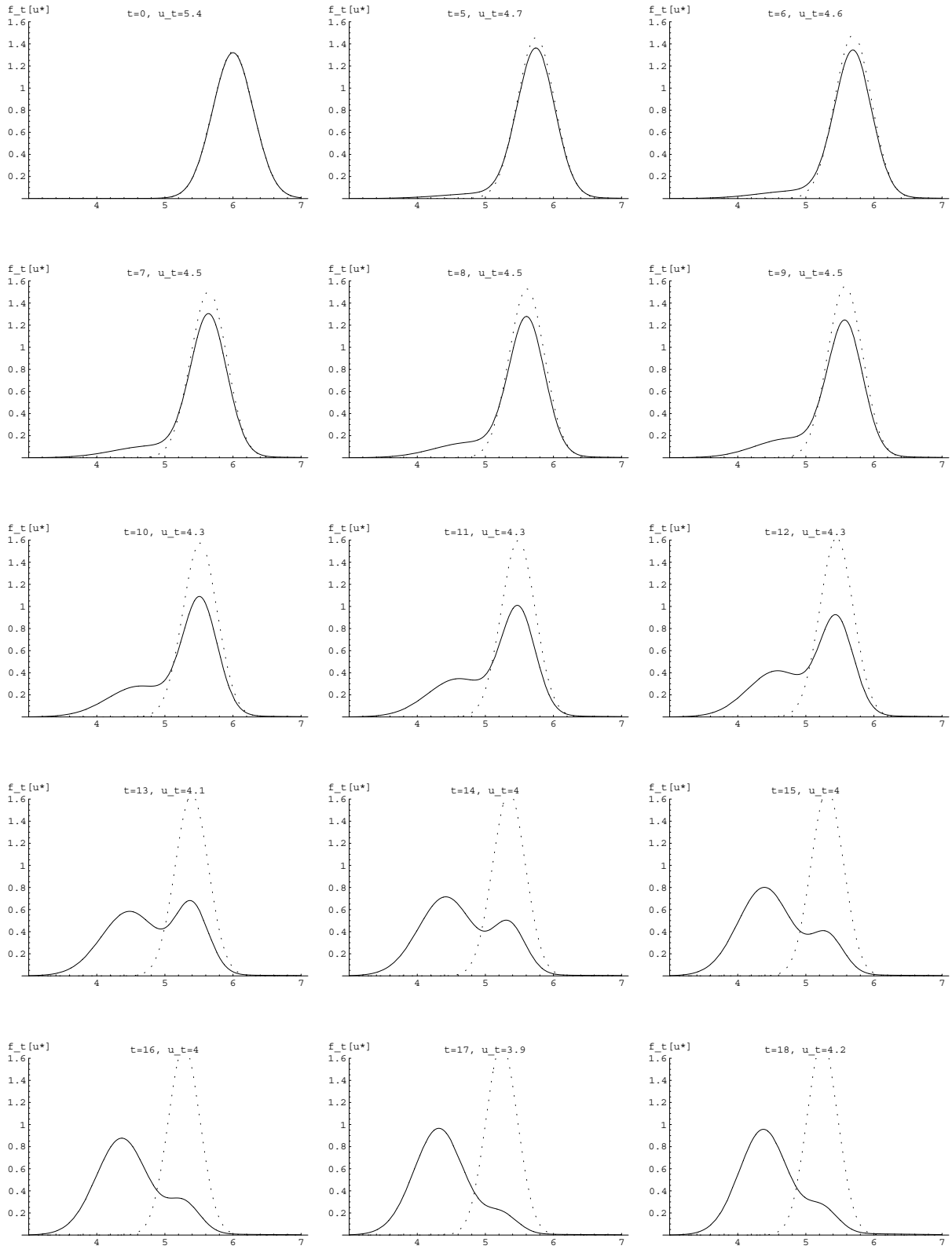
Figure 3 presents the results from simulating this model forward for 18 quarters. The solid line in each panel plots policymakers' beliefs in each period for the model described above, while the dotted line plots results for the same model *without* structural change (i.e., setting $p = 0$) for comparison. The time 0 prior for each model is plotted in the upper left-most panel; the results for the first four quarters ($t = 1, \ldots, 4$) are not reported in the figure because they are not very different from the $t = 0$ and $t = 5$ cases. Simulating the model forward requires drawing values for u_t each period; we use the quarterly values of unemployment realized in the U.S. from 1997Q1 through 2001Q1, which are listed at the top of each panel.[9]

The simulation becomes interesting after about 5 or 6 quarters. By this point, repeated realizations of u_t below policymakers' prior mean of 6 cause them to place noticeable

[8] We estimate these parameters over the period 1964–2002. We HP-filter u_t to measure u^*. The unemployment gap fits an AR(2) process better than an AR(1), but to keep the simulation down to a single state variable (u_{t-1}), we just use the dominant root of the AR(2) polynomial.

[9] Realized unemployment rates for $t = 1, \ldots, 4$ are 5.2, 5.2, 5.1, 4.9, respectively.

Figure 3: Dynamic Simulation of Policymakers' Beliefs



probability on a structural break having occurred. This is seen as a fat left tail of the distribution for u^*, which shows that beliefs about u^* are becoming increasingly diffuse. After about 11 or 12 quarters, policymakers' beliefs display a clear bimodal pattern, with "New Economy" (low u^*) and "Old Economy" (high u^*) regimes. Once policymakers' beliefs reach this point, they are extremely uncertain about the true value of u^* over a wide range of intermediate values, and they begin to revise their beliefs very strongly as new information comes in. This uncertainty persists for several quarters, and is still evident 18 quarters from the start of the simulation in the bottom right panel.

By contrast, in the model without structural change (dotted lines in Figure 3), policymakers' beliefs have a Gaussian distribution in every period. Even though policymakers update their beliefs each period in response to incoming data, the Gaussian prior and Gaussian shocks ensure that only the mean of the distribution is revised (with the variance gradually shrinking over time), while the overall Gaussian functional form is preserved.

Figure 4 maps these sequences of beliefs into policymakers' optimal estimate of the unemployment gap over time (left panel) and with respect to the observed unemployment rate (right panel).

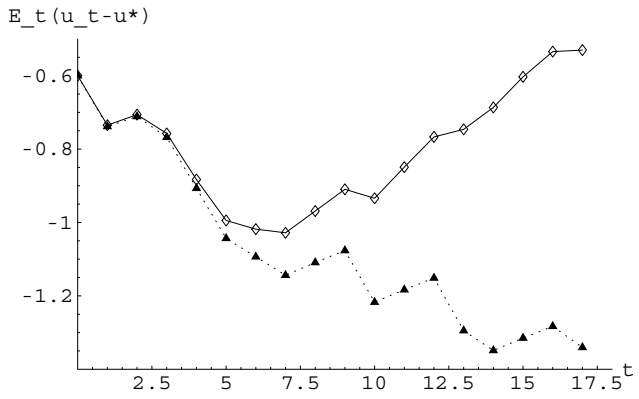


Figure 4a

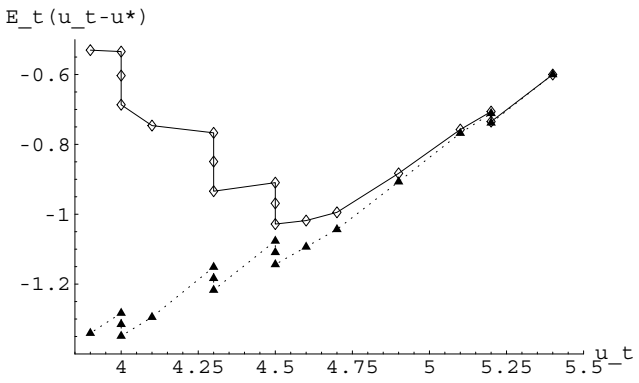


Figure 4b

The solid (dashed) lines plot policymakers' estimates in the model with (without) structural change. In the model without structural change, policymakers revise their estimates of the unemployment gap essentially linearly with the observed unemployment rate, the only deviation from linearity arising from the gradual shrinking of the variance of policymakers' beliefs about u^* over time. By contrast, the model with structural change begins

to show very strong departures from linearity after about 6 or 7 quarters—at this point, policymakers are starting to place substantial weight on the possibility that structural change has occurred, and that they may have been overestimating the size of the unemployment gap. Thus, their optimal estimate of the gap (in this simulation) is actually backward-bending, since they begin to revise their estimate of the gap backward (revising u^* at the margin even more rapidly than u_t is falling) to make up for possible previous overestimates. Although the backward-bending feature of Figure 4 depends on the specifics of the model and the simulation, the basic intuition underlying policymakers' estimates of u^* and the unemployment gap is identical to the simple model in the previous section.

Note that an interesting feature of this model of structural change is that the usual state of the world is one in which policymakers' beliefs are close to normal—and thus that optimal updating and optimal policy are usually close to linear. The nonlinearity emphasized here and in the previous section only becomes relevant when policymakers beliefs about u^* become sufficiently diffuse in a region around the mean—in this case, when policymakers begin to suspect that a structural break in the model might have occurred. This nonlinearity persists for several quarters and then gradually dies out as policymakers' beliefs converge gradually back toward normality once again.

3.2 Multivariate Signal Extraction

For simplicity, the previous sections have considered stylized models with only a single indicator variable—univariate signal extraction. In this section, we reinstate the inflation equation (1b) of the baseline model, and let policymakers update their beliefs about the unobserved variable u^* in response to both indicators: unemployment and inflation. For this more general case, policymakers' optimal inference and response functions are no longer functions simply of u_t, but now also depend on the realized inflation rate π_t, resulting in graphs that are no longer two-dimensional, as in section 2, but are instead surface plots of policymakers' optimal estimates of u^* and $u_t - u^*$ as functions of the surprises in unemployment and inflation both, as in Figures 5a and b below.

To graph these functions, we roughly calibrate the parameters of the model to the U.S. economy from 1965–2000, this time to annual data since we will focus on a single

period t rather than trace the dynamics of the model over time. These parameter values are set as follows: $\theta = 0.4$, $\alpha = 0.15$, $\beta = 0.4$, $r^* = 2.5$, $\sigma_\varepsilon = 0.5$, and $\sigma_\nu = 0.6$. We allow for a correlation between ε_t and ν_t of -0.5, also in line with annual U.S. data over this period. We initialize the model with values of $\pi_{t-1} = 2.0$ percent, $u_{t-1} = 4.2$, and $r_{t-1} = 3.0$, which correspond to core PCE inflation, unemployment, and the real federal funds rate in the U.S. in 1999. We use the uniform prior on u^* from Figure 1, which has $E_{t-1}u^* = 5$ percent. From model (1), this implies prior forecasts for unemployment u_t in 2000 of 4.76 percent and inflation π_t of 2.32 percent, respectively.

When data on unemployment and inflation for 2000 come in, policymakers in the model revise their beliefs about u^* and the unemployment gap $u_t - u^*$ on the basis of these observations as depicted in Figure 5, below:

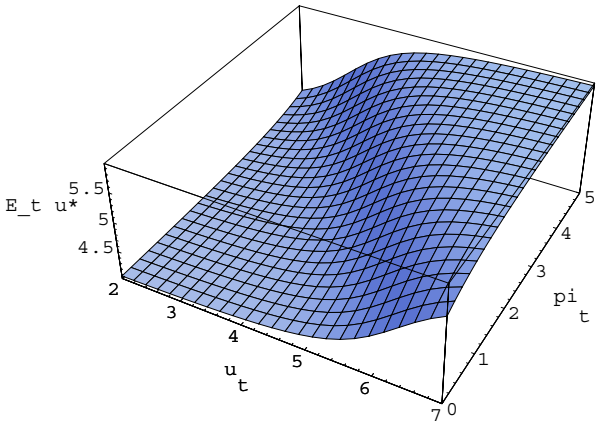


Figure 5a

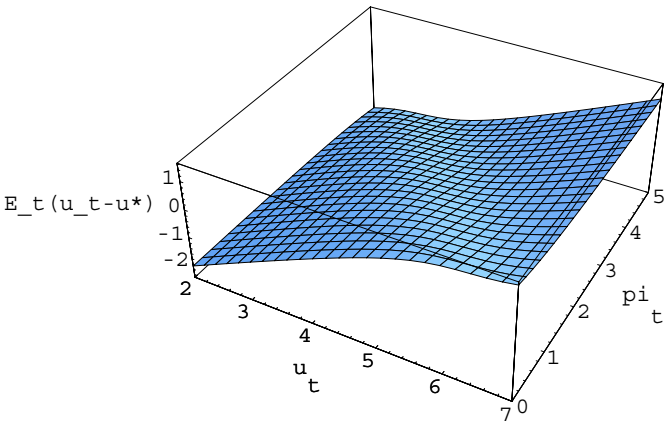


Figure 5b

In these figures, we can consider simultaneous surprises in unemployment and inflation of any magnitudes, and their effects on policymakers' optimal estimates.

The same intuition as in the baseline model holds in Figure 5: as can be seen in the right-hand panel, policymakers will update their estimate of the unemployment gap (and hence set r_t) very cautiously in response to small surprises in unemployment and inflation, where the "size of the surprise" here is the size of the joint surprise in unemployment and inflation both, as measured from the bivariate distribution that describes the shocks to these variables (in this example, a bivariate normal distribution). As the size of the surprise gets larger, policymakers will set r_t more aggressively at the margin, as they become less willing to revise their estimates of u^* further.

3.3 Forward-Looking Models

The results of this paper also apply more broadly than to just the simple backward-looking model (1) considered so far. The same ideas of signal extraction and optimal updating with a non-normal prior apply just as easily to any model that exhibits separation of estimation and control, including forward-looking models that possess this property. For example, Pearlman, Currie, and Levine (1986), Pearlman (1992), and Svensson and Woodford (2003) show that the forward-looking LQ model possesses this property. Thus, if we consider a forward-looking version of model (1):

$$(u_t - u^*) \;=\; \theta_1(u_{t-1} - u^*) + \theta_2\, E_t(u_{t+1} - u^*) + \alpha(r_{t-1} - r^*) + \varepsilon_t \tag{6a}$$

$$(\pi_t - \pi*) \;=\; \phi_1(\pi_{t-1} - \pi^*) + \phi_2 E_t(\pi_{t+1} - \pi^*) - \beta(u_{t-1} - u^*) + \nu_t \tag{6b}$$

and policymakers' preferences and information set are given, as before, by:

$$\min\,(1-\delta)\,E_t \sum_{s=t}^{\infty} \delta^{s-t}\left[(\pi_s - \pi^*)^2 + \gamma(u_s - u^*)^2\right]$$

$$\mathcal{I}_t \equiv \{\alpha, \beta, \gamma, \delta, \theta_1, \theta_2, \phi_1, \phi_2, \sigma_\varepsilon^2, \sigma_\nu^2, F_{u^*|0}(\cdot), \pi^*, r^*, \pi_t, u_t, \pi_s, r_s, u_s \mid s < t\}$$

then the above authors show that the optimal policy under commitment satisfies:[10]

$$r_t - r^* \;=\; c\,E_t(u_t - u^*) + d\,(\pi_t - \pi^*) \tag{7}$$

where c and d are constants that can be computed from the parameters θ_1, θ_2, ϕ_1, ϕ_2, α, β, γ, and δ using the methods described in those papers.[11] The analysis of optimal policy with a non-normal prior is then exactly the same as before: policymakers form an optimal estimate of the unemployment gap, $E_t(u_t - u^*)$, and then set policy according to this estimate. So long as the control stage of the problem is separate from that of estimation, the same types of diagrams presented earlier remain relevant for thinking about the problem, and the same intuition applies.

[10] The optimal policy under discretion is also linear, although the coefficients c and d would be different.

[11] Note that the property of separability of estimation and control in equation (7) is a special feature of the LQ model under optimal policy. For example, if one considers a *simple* rule rather than the fully optimal rule, then estimation and control are not separable and the solution to the model becomes much more difficult computationally (Meyer, Swanson, and Wieland (2001) consider this case). Similarly, if any of the parameters of the model are not known with certainty, then the optimal policy must take into account learning about these parameters by policymakers, which again becomes computationally much more difficult (Wieland (2003) analyzes this case).

Figure 6a

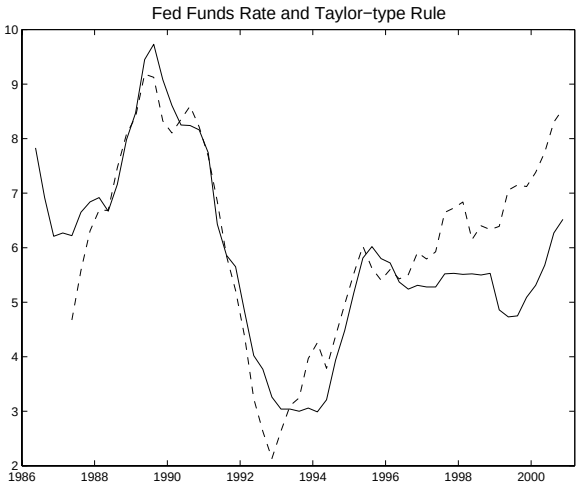


Figure 6b

4. Discussion

For the diffuse-middled class of prior distributions analyzed above and in the Appendix, it is optimal for policy to respond cautiously to small surprises in the observed unemployment rate, but to respond increasingly more aggressively at the margin as the size of the surprise gets larger. These features of optimal policy match statements by Federal Reserve officials in the late 1990s presented in the Introduction, and Figure 6 suggests that it also matches the behavior of the Fed over this period.

The left-hand panel, Figure 6a, plots the quarterly U.S. unemployment rate from 1986 to 2001, along with a dashed horizontal line at $u_t = 6$ to depict a common estimate of the natural unemployment rate in the mid-1990s (e.g., Staiger, Stock, and Watson (1997)). The right-hand panel, Figure 6b, compares the Federal Reserve's actual interest rate policy (solid line) to what would have been prescribed by a Taylor (1993) type rule (dashed line) based on GDP deflator inflation (not shown) and the unemployment rate data in Figure 6a.[12]

The Taylor-type rule in Figure 6b tracks the Fed's actual behavior fairly well until about 1997, at which point the two diverge noticeably. In particular, the Fed appears to have been much less aggressive between 1997 and 1999 than a Taylor rule based on

[12] The Taylor-type rule in Figure 6b is given by $i_t = r^* + \pi_t + \alpha(\pi_t - \pi^*) - \beta(u_t - u^*)$, where $r^* = 2.75$, $\pi^* = 2$, $u^* = 6$, $\alpha = 0.5$, and $\beta = 1.8$, u_t denotes the unemployment rate in Figure 6a, and π_t the four-quarter change in GDP deflator inflation.

previous estimates of u^* would have implied. Beginning sometime in 1999, however, the Fed appears to have returned to the marginal responsiveness implied by the Taylor Rule—i.e., the marginal prescriptions from the Taylor rule seem to coincide with the Fed's marginal interest rate moves in 1999 and 2000, even though the levels had diverged as a result of the Fed's earlier cautiousness. The Fed's behavior over this recent historical episode—an attenuation in the responsiveness of policy for a time, followed by a return to a more aggressive response at the margin—seems to match the intuition behind optimal updating and optimal policy presented in this paper.

5. Conclusions

The literature on optimal monetary policy typically makes three major assumptions for tractability and simplicity: 1) policymakers' preferences are quadratic, 2) the economy is linear, and 3) stochastic shocks and policymakers' priors about unobserved variables are normally distributed. The linearity of the optimal policy that results from these assumptions is an advantage when maximum simplicity is required, but arguably detracts from the realism of the model in many cases.

This paper argues that relaxing the third assumption provides a tractable and realistic model of nonlinearities in the optimal and actual conduct of policy. Tractability is maintained through the principle of separation of estimation and control, which continues to hold in the LQ framework despite the non-normal priors that are the focus of this paper. For example, separability of estimation and control allows the basic insights of the paper to be extended in a natural way to forward-looking LQ models, which have become a baseline for much of the recent literature on optimal monetary policy.

The empirical relevance of the nonlinearities presented in this paper are demonstrated in two ways. First, a simple simulation shows how "diffuse-middled" priors can arise naturally in a dynamic model with structural change. Second, statements by Federal Reserve officials and the behavior of the Fed in the 1990s are shown to line up with the basic insights of the model. Thus, the optimal nonlinearities demonstrated in this paper appear to be relevant in practice as well as in principle.

Finally, the results demonstrated in this paper are not specific to models of monetary policy comprising an output equation and an inflation equation, but rather apply to any signal extraction problem with non-normal priors on a key unobserved variable of the model. Since policymakers in a variety of contexts face the problem of updating priors about unobserved variables through signal extraction, and these priors may well be more diffuse in a region around the mean than a Gaussian distribution, the basic insights in this paper should be regarded as being potentially applicable to a wide variety of situations.

Appendix: Mathematical Derivations

Let policymakers' prior (time $t-1$) distribution on u^* be given by the density function $f(u^*)$. The mean of this distribution is assumed to exist, and without loss of generality, equals 0. The variance is assumed to exist and equals σ^{*2}. Given u^*, the (observable) unemployment rate u_t is distributed $N(u^*, \sigma^2)$, with density denoted by ϕ. We now show what qualities of the density f lead to the features of optimal nonlinear policy emphasized in the text, namely: (1) that the responsiveness of policy is attenuated relative to normal for small surprises ($u_t \approx 0$), and (2) that the marginal responsiveness of policy increases in the size of the surprise in u_t.

We restrict attention to the case of one observable variable, u_t, for clarity. The analysis for multiple observable variables is essentially identical—one need only calculate the size of the joint surprise to determine the extent of updating of policymakers' priors that takes place.

The formula for policymakers' posterior on u^*, $E_t u^* \equiv E[u^*|u_t]$, is the usual:

$$E[u^*|u_t] = \frac{\int u^* \phi(u_t - u^*) f(u^*) du^*}{\int \phi(u_t - u^*) f(u^*) du^*} \tag{A1}$$

We will make use of the following Lemma:

Lemma 1: Assume that policymakers set the interest rate r_t as a decreasing linear function of the estimated unemployment gap, $u_t - E[u^*|u_t]$. Then the marginal responsiveness of policy to observed unemployment, $\partial r_t / \partial u_t$, is proportional to:

$$\frac{\partial E[u^*|u_t]}{\partial u_t} - 1 \;=\; \frac{1}{\sigma^2}\mathrm{Var}[u^*|u_t] - 1 \tag{A2}$$

the convexity of policy, $\partial^2 r_t / \partial u_t^2$, is proportional to:

$$\frac{\partial^2 E[u^*|u_t]}{\partial u_t^2} \;=\; \frac{1}{\sigma^4}\,\mathrm{Skew}[u^*|u_t] \tag{A3}$$

and $\partial^3 r_t / \partial u_t^3$, is proportional to:

$$\frac{\partial^3 E[u^*|u_t]}{\partial u_t^3} \;=\; \frac{1}{\sigma^6}\,\text{Exc. Kurt}\,[u^*|u_t] \tag{A4}$$

where $\text{Exc. Kurt}[u^*|u_t] \equiv E[(u^* - \mu)^4|u_t] - 3\big(\mathrm{Var}[u^*|u_t]\big)^2$ and $\mu \equiv E[u^*|u_t]$.

Proof: (A2)–(A4) follow from (A1) by successive differentiation through the integral. □

Note that for f normally distributed, $\mathrm{Var}[u^*|u_t]$ is independent of u_t and less than 1, so that the marginal responsiveness of policy to u_t is constant and negative.

By equation (A2), property (1) above is equivalent to $\text{Var}[u^*|u_t]$ being larger than normal when evaluated at $u_t = 0$.[13] We begin with the assumption that f is Gaussian, and consider what perturbations increase the conditional variance in (A2). This is essentially a calculus of variations problem. The following diagram serves as a useful illustration:

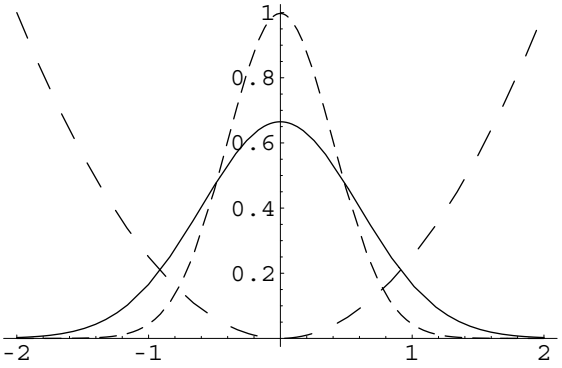


Figure A1

where the solid line in the figure depicts the density f, the short-dashed line ϕ, and the long-dashed line the function $y = x^2$. The quantity $\text{Var}[u^*|u_t = 0]$ is (up to a constant factor) the integral of the product of these three functions:

$$\text{Var}[u^*|u_t=0] = \frac{\int u^{*2}\phi(-u^*)f(u^*)du^*}{\int \phi(-u^*)f(u^*)du^*} \tag{A5}$$

Holding the denominator constant, we can think of increasing the value of this quantity by reducing f where the product $u^{*2}\phi(u^*)$ is small, and increasing f where the same product is large. Thus, we reduce f near 0, and increase f for intermediate values of u^*.

We prove this formally via a variational argument on f as follows:

Proposition 1: For prior distributions f_δ that are sufficiently close to a Gaussian distribution f, property (1) (policy attenuation near 0) is possessed by distributions f_δ that are more diffuse than f in a region around the mean.

Proof: Consider a perturbation of the density function f by an amount δ over the interval $[\xi-h, \xi+h]$, and call this new function f_δ. We have:

$$\begin{aligned} \frac{\partial}{\partial\delta}\int u^{*2}\phi(u^*)f_\delta(u^*)du^* &= \frac{1}{\delta}\int_{\xi-h}^{\xi+h} u^{*2}\phi(u^*)\delta du^* \\ &\approx 2h\xi^2\phi(\xi) \end{aligned} \tag{A6}$$

[13] There is nothing that prevents $\text{Var}[u^*|u_t] > 1$ and (A2) from being positive (e.g., in a two-regime model such as in section 3 or Swanson (2000)). In such a situation, we interpret a further increase in (A2) (toward infinity) as the desired policy attenuation.

Similarly,

$$\frac{\partial}{\partial \delta} \int \phi(u^*) f_\delta(u^*) du^* = \frac{1}{\delta} \int_{\xi-h}^{\xi+h} \phi(u^*) \delta du^* \approx 2h\, \phi(\xi) \tag{A7}$$

The derivative of (A5) with respect to the δ-perturbation is then:

$$\frac{2h\xi^2 \phi(\xi) \int \phi(u^*) f(u^*) du^* - 2h\, \phi(\xi) \int u^{*2} \phi(u^*) f(u^*) du^*}{\Big(\int \phi(u^*) f(u^*) du^* \Big)^2} \tag{A8}$$

This quantity is less than 0 for $\xi = 0$. It is actually *greater* than 0 as $|\xi|$ tends toward infinity, though the effect is very small (because $\phi(\xi)$ is so small). More specifically, it is negative for $\xi^2 < E[u^{*2} | u_t = 0]$ and positive for $\xi^2 > E[u^{*2} | u_t = 0]$. Thus, perturbing f downward at 0 (and pushing it upward for intermediate and large values of $|\xi|$) leads policy to have property (1) emphasized in the text. □

Note that equation (A8) holds for non-Gaussian base distributions f as well, so that the variational argument is valid for f significantly different from a normal distribution, as well as very close to Gaussian.

We can use the same technique to determine what distributions f possess property (2). We'll make use of the following lemma:

Lemma 2: Let f be a density function and let f_δ be a perturbation of f by an amount δ over the interval $[\xi - h, \xi + h]$. Let

$$\mu_{\delta,n} \equiv \frac{\int u^{*n} \phi(-u^*) f_\delta(u^*) du^*}{\int \phi(-u^*) f_\delta(u^*) du^*} \tag{A9}$$

be the nth moment of f_δ conditional on $u_t = 0$. Then

$$\frac{\partial \mu_{\delta,n}}{\partial \delta} = \frac{2h\phi(\xi)}{\int \phi(u^*) f(u^*) du^*} (\xi^n - \mu_{0,n}) + o(h) \tag{A10}$$

where $\lim_{h \to 0} o(h)/h = 0$.

Proof: Simple generalization of equations (A6)–(A8). □

Proposition 2: For prior distributions f_δ that are sufficiently close to a Gaussian distribution f, property (2) (an increasingly aggressive response to unemployment at the margin) is possessed for u_t near 0 by distributions f_δ that are more diffuse than f in a region around the mean and with tails that are thinner than f.

Proof: Property (2) states that policy should be concave for $u_t > 0$ and convex for $u_t < 0$. From equations (A3) and (A4), this is equivalent to Skew$[u^*|u_t=0] = 0$ and Exc. Kurt$[u^*|u_t=0] < 0$. We wish to know for which distributions f_δ this is the case.

A straightforward calculation yields:

$$\text{Exc. Kurt}[u^*|u_t=0] \;=\; \mu_{\delta,4} - 4\mu_{\delta,3}\mu_{\delta,1} - 3\mu_{\delta,2}^2 + 12\mu_{\delta,2}\mu_{\delta,1}^2 - 6\mu_{\delta,1}^4 \tag{A11}$$

where $\mu_{\delta,n}$ is the nth moment of f_δ as defined in (A9). Differentiating (A11) with respect to δ, evaluating at $\delta = 0$, and noting that $\mu_{0,1} = 0$ for the base Gaussian density f yields:

$$\frac{\partial}{\partial\delta}\big(\text{Exc. Kurt}[u^*|u_t=0]\big) \;=\; \mu'_{\delta,4} - 4\mu_3\mu'_{\delta,1} - 6\mu_2\mu'_{\delta,2} \tag{A12}$$

where $\mu_n \equiv \mu_{0,n}$ and a prime denotes differentiation with respect to δ. Using Lemma 2 yields:

$$\frac{\partial}{\partial\delta}\big(\text{Exc. Kurt}[u^*|u_t=0]\big) \;\approx\; \frac{2h\phi(\xi)}{\int\phi(u^*)f(u^*)du^*}\left[\xi^4 - \mu_4 - 4\mu_3(\xi-\mu_1) - 6\mu_2(\xi^2-\mu_2)\right] \tag{A13}$$

which simplifies to:

$$\frac{2h\phi(\xi)}{\int\phi(u^*)f(u^*)du^*}\left[\xi^4 - 6\mu_2\xi^2 + 3\mu_2^2\right] \tag{A14}$$

after using $\mu_1 = 0$, $\mu_3 = 0$, and $\mu_4 = 3\mu_2^2$ for the base Gaussian density f.

Now, the polynomial in ξ inside the square brackets in (A14) is positive at $\xi = 0$, is positive as $|\xi| \to \infty$ (though the effect on (A14) is small because $\phi(\xi)$ is small), and has four real roots at $\xi = \pm\sqrt{(3 \pm \sqrt{6})\,\mu_2}$, hence is negative for intermediate values of ξ—specifically, those values for which $|\xi| \in \left[\sqrt{(3-\sqrt{6})\mu_2}, \sqrt{(3+\sqrt{6})\mu_2}\,\right]$. Thus, property (2) is satisfied by distributions f_δ that are perturbed downward at 0, downward in the tails, and upward for intermediate values of ξ. Keeping the density f_δ symmetric ensures that Skew$[u^*|u_t=0] = 0$. □

Note that—in contrast to property (1)—property (2) is a feature of distributions that have *thin* tails, but both property (1) and (2) are possessed by distributions that are more diffuse in a region around the mean than a base Gaussian density f.

References

Bertsekas, Dimitri. *Dynamic Programming* (Prentice-Hall: Englewood Cliffs, NJ, 1987).

Filardo, Andrew. "New Evidence on the Output Cost of Fighting Inflation," *Federal Reserve Bank of Kansas City Economic Review* 83 (1998), pp. 33–61.

Levy, David. "Interview with Roger W. Ferguson, Jr." *The Region* 14 (June 2000), Federal Reserve Bank of Minneapolis.

Meyer, Laurence. "Q & A on the Economic Outlook and Challenges Facing Monetary Policy," speech before the Philadelphia Council for Business Economics, Federal Bank of Philadelphia, September 8, 1999; transcript available from the Federal Reserve Board's public web site.

Meyer, Laurence, Eric Swanson, and Volker Wieland. "NAIRU Uncertainty and Nonlinear Policy Rules," *American Economic Review* 91 (2001), pp. 226–31.

Orphanides, Athanasios and Volker Wieland. "Inflation Zone Targeting," *European Economic Review* 44 (2000), pp. 1351–87.

Pearlman, Joseph. "Reputational and Nonreputational Policies under Partial Information," *Journal of Economic Dynamics and Control* 16 (1992), pp. 339–57.

Pearlman, Joseph, David Currie, and Paul Levine. "Rational Expectations Models with Partial Information," *Economic Modelling* 3 (1986), pp. 90-105.

Sack, Brian and Volker Wieland. "Interest-Rate Smoothing and Optimal Monetary Policy: A Review of Recent Empirical Evidence," *Journal of Economics and Business* 53 (2000), pp. 205-28.

Staiger, Douglas, James Stock, and Mark Watson. "How Precise Are Estimates of the Natural Rate of Unemployment?" in *Reducing Inflation: Motivation and Strategy*, ed. Christina Romer and David Romer (Univ. of Chicago Press: Chicago, 1997), pp. 195–242.

Svensson, Lars and Michael Woodford. "Indicator Variables for Optimal Policy," *Journal of Monetary Economics* 50 (2003), pp. 691–720.

Swanson, Eric. "Optimal Nonlinear Policy: Signal Extraction with a Non-Normal Prior," unpublished manuscript, Federal Reserve Board (2000).

Swanson, Eric. "On Signal Extraction and Non-Certainty-Equivalence in Optimal Monetary Policy Rules," *Macroeconomic Dynamics* 8 (2004), pp. 27–50.

Taylor, John. "Discretion versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy* 39 (1993), pp. 195–214.

Taylor, John. "A Historical Analysis of Monetary Policy Rules," in *Monetary Policy Rules,* ed. John Taylor (Univ. of Chicago Press: Chicago, 1999), pp. 319–41.

Wieland, Volker. "Monetary Policy and Uncertainty About the Natural Unemployment Rate," *CEPR Discussion Paper* 3811 (2003).