STICKINESS

CHRISTOPHER A. SIMS

Abstract. We discuss an array of models of dynamically optimizing representative firms and workers, with inertia and price-wage stickiness modeled in various ways. The degree of price and wage stickiness bears no necessary connection to the strength of real effects of monetary policy. Matching the combination of real and nominal inertia in responses to monetary policy found in the data requires a more complex model, with more sources of stickiness and inertia, than has been standard in the literature. The pervasiveness of sluggish cross-variable responses in the macro data, combined with the implausibility of many of the microeconomic stories underlying adjustment cost models, suggests that we look for a different approach to modeling the sources of inertia in both prices and real variables. One such approach, based on limited information-processing capacity, is sketched.

I. Introduction

Most prices and wages do not change daily, or even weekly, and many seldom change within a year. Keynes argued that this observation played a central role in an explanation of how an economy could operate for long periods of time at inefficiently low levels of activity and that it implied strong effects of monetary and fiscal policy on the level of economic activity. He gave little attention to the microeconomic foundations of wage stickiness, but proceeded to incorporate it directly into an equilibrium model of the economy as a whole, constructed as a collection of behavioral equations. The basic conclusion that wage and price stickiness is a source of non-neutrality in the effects of changes in nominal aggregate demand[1] is accepted even by many economists who do not agree with Keynes's diagnosis or prescriptions for aggregate fluctuations.

This paper criticizes the kinds of models of price and wage stickiness now common in the literature from several directions. It begins with a discussion of the foundations of the theory of stickiness in individual behavior. Though early work in rational expectations modeling often emphasized search behavior and signal extraction problems as

Date: April 2, 1999.

[1] I do not myself find "aggregate demand" to be a very useful, or even a very clearly definable, notion, but here I am trying to summarize conventional views.

a source of an observed Phillips Curve[2], more recent theory has not emphasized these elements. It may be that this reflects both the difficulty of modeling information-processing rigorously and the fact that doing so forces us to confront the limitations of optimizing theories of behavior. It does not seem to reflect any empirical evidence that information-processing is not important to stickiness phenomena. Indeed there is a great deal of informal evidence that it is indeed important to stickiness. Section II elaborates this point and argues that it is important to policy conclusions that we return our attention to the difficult task of modeling information processing and its consequences in macro models.

Empirical evidence on the joint dynamic behavior of prices, wages, and macroeconomic aggregates shows certain strong regularities that do not fit easily into any existing macroeconomic theories. Section III lays out these facts and the difficulties they pose for existing modeling approaches. Existing models tend to be either "classical", in the sense that they imply that most kinds of disturbance to the economy should result in instant movements in prices, or "Keynesian", in the sense that they imply that most kinds of disturbances in the economy should produce instant movements in output. The data show neither pattern.

The next several sections examine the behavior of a variety of models that differ mainly in how they model real and nominal stickiness. The models are laid out in a common format, and are solved via local linear approximations around their steady states. They are formulated in continuous time to avoid the need to use the uninterpretable "one period" delays that plague the discrete time models in this literature. None of them, unfortunately, meets the criterion of being based on explicit recognition of the importance of information-processing. They are meant to display the differences and similarities in predictions about aggregate behavior across models of various existing types. They are meant also to suggest directions in which existing modeling approaches might be modified to fit more closely with the data and the insights available from informal thinking about the role of information processing.

It is of course widely recognized that wages do not behave like a spot price in an instantaneously clearing market, and economists of a classical orientation[3] have long argued that wage data, not having allocative significance, can legitimately be left out of macroeconomic models. It is, however, possible to introduce wages that have the dynamics of Taylor or Calvo style overlapping contracts, that guide continuously clearing markets, and that nonetheless imply the kind of short and long run neutrality of nominal policy interventions that is associated with classical models. Models of this type, combined with assumptions on technology that keep real adjustments to

[2]Robert E. Lucas (1973), Phelps (1968), e.g.

[3]Hall and Lilien (1979), Barro (1979), e.g.

shocks sluggish, provide one possible route toward matching the empirical facts. The nature of these models is laid out in section IV. When these models' sticky price behavior is combined with technology-based inertia in real variables, they may come close to matching the qualitative facts of observed aggregate dynamic price-quantity relationships. With realistic specifications of the real role of money, such models imply negligible real effects of fiscal and monetary policy; there is evidence in the data for such effects, albeit not overwhelming evidence.

The Keynesian idea is to postulate sticky wage and price adjustment and also that quantities react promptly to interest rates, taxes, and the slowly changing price signals. Keynes himself stressed the importance of expectations, and even in places of the endogeneity of expectations. He did not account for them explicitly in his mathematics, however, and they vanish altogether when his thinking is reduced to the standard ISLM abstraction. Sticky wages and prices, combined with optimizing, forward-looking behavior by agents, generate powerful effects of monetary policy. Clear understanding of these effects and how they operate requires modeling the intertemporal government budget constraint explicitly. Two Keynesian models are presented in section V.2. The first specifies ad hoc price and wage dynamics based on the gaps in workers' and firms' first-order conditions with respect to labor. This amounts to appending a "Phillips Curve" and a "Markup Equation" to the model. The second model adopts a New Keynesian setup, in which monopolistically competitive firms set prices subject to adjustment costs, and monopolistically competitive workers set wages subject to adjustment costs. The two models produce qualitatively similar results.

The Keynesian models make quantities counterfactually responsive to policy and to other disturbances, however. Section V.4 presents two models, one New Keynesian and one with "ad hoc" price adjustment, that keep employment smooth by giving it adjustment costs and place adjustment costs on bond-holdings, so that sharp fluctuations in taxes or income tend to show up initially in money accumulation by workers. These models can produce apparently realistic qualitative behavior.

My own view is that the best of these models ought to be regarded as proxies for better, harder-to-construct models that would recognize the implications of limited information-processing capacity. Recognizing this suggests reasons for skepticism in application of these models to drastic policy experiments. Precisely those models that have been constructed to avoid the "Lucas critique" are then subject to a set of caveats that may sound much like the Lucas critique itself. These points are laid out in the concluding section VII.

II. Stickiness as a Consequence of Limited Information-Processing Capacity

Sticky prices are most common in markets with highly heterogeneous products and in which on at least one side of the market agents are only occasional purchasers. The labor and housing markets are good examples, as are most consumer goods markets. There is a plausible explanation for this. Heterogeneity of products makes assessing the average price level in the market a statistically demanding task, and assessing the relevance of the average price to an individual seller or buyer's market situation is an additional nontrivial task. Because individuals have many things to think about and limited time, they can devote only limited intellectual resources to these tasks of data-gathering and analysis. Individuals setting prices therefore do not change them frequently, or in response to every change in market conditions. This reflects both the fact that the price setters don't have time to make prices respond instantly to every disturbance, and the fact that those searching across prices will find it more convenient to deal with price-setters whose price quotes have stability over time.

Early work on rational expectations emphasized the relationship of search theory to unemployment. Robert E. Lucas (1973) motivated a delayed reaction by workers to changes in the price level by suggesting they faced a signal-extraction problem, in which it was difficult to determine the aggregate price level in the short run. Since data on the money stock and other instruments of aggregate policy are available with little delay, though, it seems implausible to postulate a technological barrier to individuals determining them precisely within a short time, and this makes it difficult to explain why they should react with a delay to policy measures. This kind of reasoning may explain why signal-extraction models are not common in the recent macroeconomic literature.

But the intuitive appeal of the Lucas signal-extraction story does not really rest on whether an aggregate price index is published with little delay. We know from personal experience that many data that we could look up daily in the financial press, and that are in principle relevant to our optimal economic decision-making, do not in fact influence our behavior, except when they change dramatically, or perhaps when we occasionally set aside some time to reassess our portfolio. So even though the money stock, the Federal Funds rate, and the CPI are regularly published, we know that it is plausible that people with limited time, aspects of their lives other than economics to attend to, and many important micro-economic signals relevant to their personal situation to react to, do not generally react to aggregate signals instantly as soon as they are publicly known.

Information theory, which formally models physical limits to the rate of transfer of information, may provide a way for us to capture the intuitive appeal of the signal-extraction story while neither introducing implausible technological constraints on the observability of data nor abandoning entirely the strategy of modeling behavior as reflecting optimization. Individual behavior can be modeled as reflecting optimal patterns of response to market (and other) signals, given the limited information-processing capacity of the individual.

In this paper we do not provide any example of a model that is formally based on such information-theoretic analysis. The Appendix provides an introduction to information theory and explains two of its important implications for the form of dynamic responses by individuals to market signals. If the response is related to the signal by a finite-capacity channel,

i. there must be a random, or noise component to the response, with the noise component dominating the systematic component at high frequencies, and
ii. the response must be smoothed, and hence delayed, relative to the signal.

These qualitative results seem to fit well with what shows up in the aggregate data. There are therefore theoretical as well as empirical reasons to prefer models that imply delayed responses to market signals for any variables that reflect the aggregate behavior of individuals, each separately required to pay attention to the signal.

In sections V.4 and V.1 below we present models that have the qualitative characteristics that we have argued emerge from treating individuals as having finite information-processing capacity. Models derived as these are, from individual, continuous-time optimizing behavior with pervasive adjustment costs[4] may approximate well behavior that is actually based on limited information-processing capacity. But in some respects, particularly when we consider using the model to project effects of permanent changes in the stochastic specification for policy, they are likely to be misleading. Like most standard rational expectations models, these models make the distinction between anticipated and unanticipated policy shifts central, whereas information-based modeling would not. Like the original Lucas signal-extraction formulation, information-based models will imply that when policy shocks have large variance, responses to policy will be more than proportionally larger, but the information-based models will imply also that delays in response to policies with large variation will be shorter. And of course there is the central distinction, that signal-extraction and rational expectations models suggest that providing complete information about policy

[4]The overlapping contract structure in section V.1 is not explicitly based on adjustment costs. However, it seems natural to interpret this form of agreement, in which each worker sets a price and a rate of effort that stays fixed over some span of time, as in part reflecting aversion to information-processing costs on the part of workers.

actions will have important effects on the response to policy, whereas information-based models imply that complicated information about policy may be ignored by agents who do not find it worthwhile to devote much attention to the information.

III. Ugly Facts

In a recent paper (Keating 1997) has examined the joint dynamic behavior of a variety of price and output time series, concluding that at both quarterly and monthly levels price and output innovations are very weakly related.[5] Relations between the variables tend to be confined to smooth, slow impulse responses. Keating considers only bivariate interactions, but his observations carry over to larger models. In such models we can observe further that prices, wages, and other nominal variables show fairly strong, albeit delayed, dynamic interactions. These interactions imply that prices and wages are in fact responding to the state of the economy and influencing it. Ignoring their variation is therefore difficult to justify.

These points are brought out in Figure 1, which displays results from a 7-variable reduced form vector autoregression that includes both CPI and average hourly earnings, as well as commodity prices. Because of the triangular normalization of the model, first-period responses in all the responses in the upper right triangle of graphs are by construction zero. The graphs display responses with two-standard-deviation Bayesian error bands.[6] Only a few cross-variable (i.e., off-diagonal) graphs show significantly nonzero (in the sense that zero is not in the two-σ band and a substantively important non-zero value is so included) initial period responses. The strong Okun's law connection between unemployment and output appears in the (2,1) position on the figure, and a strong relation between production and hourly earnings, almost certainly representing overtime rules, appears in the (4,1) position. Weaker, but non-negligible instantaneous responses are present connecting the interest rate to output and unemployment, in the (6,1) and (6,2) positions. Commodity prices may not seem to show important contemporaneous connections to output and CPI (the (7,1) and (7,3) positions), but the scale of variation of commodity prices is so much greater than for wages or prices that responses of the same size in log units as the contemporaneous response of wages to output (about .12) are difficult to see in the PCOMM row of graphs. Since the interest rate and commodity prices are set in continuously clearing auction markets, they are not expected to be sticky. The quick response of

[5] Keating concludes that price stickiness must play an important role in business cycle fluctuations, probably because he has in mind the polarity across existing models, in which price stickiness implies strong real responses to nominal disturbances. We will argue below that observing stickiness does not imply that real cyclical dynamics depend on the stickiness.

[6] The estimates use a Bayesian reference prior like that described in Sims and Zha (1998) and applied in Leeper, Sims, and Zha (1996) and Sims (forthcoming).

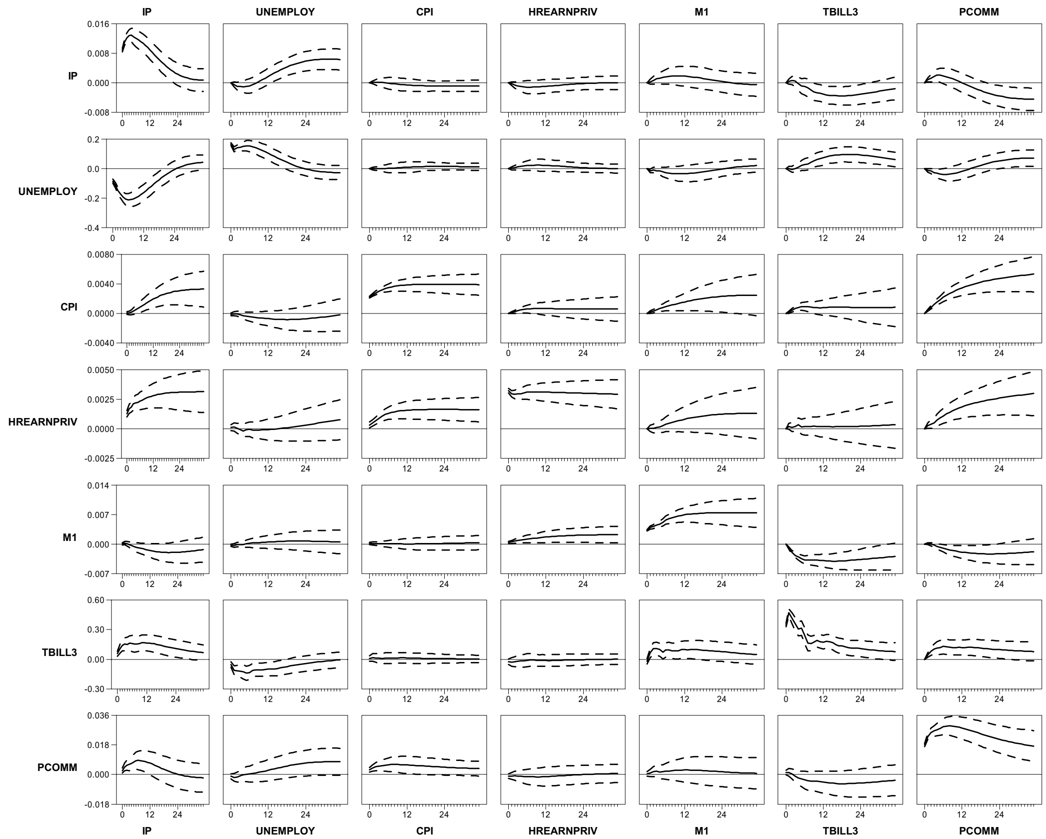

FIGURE 1. Impulse Responses for a Reduced Form

unemployment to output does not represent response of individual decision-makers to an aggregate signal, so it does not conflict with our information-theoretic interpretation. The response of wage to output would contradict our information-theoretic approach, except that it almost certainly represents the mechanical operation of overtime clauses in labor contracts, state laws, and company personnel policies. The size of the effect observed—a change of .0012 in wage when output changes by .008 in log units—is about one third what would be expected if all workers were just above 40 hour workweeks and earning a 50% overtime premium on hours over 40.

Other multivariate macro models in the literature have produced similar results. Leeper, Sims, and Zha (1996), for example, include both wages and CPI in their 18-variable model, and find almost the same patterns as in the 7-variable reduced form shown here.[7]

[7]The exception is that they do not show a strong contemporaneous connection of hourly earnings to output.

Keating himself interprets his finding of delayed response of price variables to disturbances in real activity as implying strong real effects of nominal disturbances. If we confine our attention to the range of most commonly used macroeconomic models, this conclusion would be justified. Keynesian models postulate price rigidity and imply strong real effects of nominal disturbances; market-clearing models that imply little effect on real variables of nominal disturbances imply strong effects of real disturbances on prices.

But these conclusions should not be accepted so easily. The finding of sluggish cross-variable interactions in empirical macroeconomic models is too pervasive to fit easily into any of the existing standard theoretical frameworks. There are three directions in which these models might be extended to bring them more into line with the facts. One, already discussed above, is a fundamental reorientation of modeling style, in which limited information-processing capacity of agents is recognized explicitly. While I regard this direction as the most promising in the long run, it is not pursued formally in this paper. A second is to find ways to interpret sticky prices in a market-clearing model—which we do in sections IV and V.1. And a third is to find ways to introduce sufficient real friction or inertia in Keynesian models that they can reproduce the observed qualitative nominal-real dynamic interactions. Such Keynesian models with pervasive adjustment costs, explored in section V.4, may turn out to interpretable as the outcome of optimization subject to information-processing capacity limits.

IV. Neutral Stickiness: A Lean Classical Sticky-Wage Model

Explicit price and wage contracts of at least some forms could exist without implying economic behavior of real variables any different from what would be observed with continuously clearing spot markets. While the point seems obvious once one has understood it, there may not be another model in the literature that actually makes the point. We first present a bare bones model that attempts to make the idea transparent. The model considers stickiness in wages only; it has no capital. Later, when we compare a variety of more realistic models in V, we expand the model to add money and capital, to examine its ability to produce plausible implications for the economy's behavior.

IV.1. **Representative household.**

$$\max_{C,L,\ell,B_C,M} \int_0^\infty e^{-\beta t} U(C,\, 1-L)dt \tag{1}$$

subject to

$$\lambda: \qquad C \cdot (1 + \gamma V) + \frac{\dot{B}_C + \dot{M}}{P} + \tau = \frac{rB}{P} + \frac{Y}{P} + \pi \tag{2}$$

$$\mu: \qquad \dot{Y} - w\dot{L} \leq -\delta \cdot (Y - wL) \tag{3}$$

$$B_C \geq 0, M \geq 0 \tag{4}$$

$$V = \frac{PC}{M} \,. \tag{5}$$

Here we are using a convention we will follow throughout. As each constraint is displayed, it is labeled on the left with the Lagrange multiplier that is associated with it in the first order conditions (FOC's) displayed later. When the FOC's are displayed, they are labeled on the left by "∂X", where X is the variable with respect to which the FOC has been taken.

Equations (2) and (3) describe an ordinary budget constraint with transactions costs, of the same type that has appeared in Leeper and Sims (1994) and Sims (1994), for example. What is different about this model is (3), which arises from the contracting structure. To understand where it comes from, it is perhaps easiest to see that it can be derived from the following two constraints:

$$\dot{Y} \leq w\ell - \delta Y \tag{6}$$

$$\dot{L} \geq \ell - \delta L \tag{7}$$

Here $\ell(t)$ represents the rate at which new contracts are being signed at t. The rate of growth of total labor input is this rate, less the rate at which old contracts are expiring, given by δL. The rate of growth of the wage bill is the rate of addition of new contract payment obligations, $w\ell$, less the rate of decay of old contract payment obligations, δY. The model makes sense if we use (6) and (7) explicitly in place of (3), but it makes more sense if we use (3) alone. This avoids any implication that Y and L must individually have differentiable time paths, i.e. that ℓ exists as a real-valued function of time. We may want to allow for the possibility that discrete lumps of contracts are signed at a given date, or in a stochastic version of the model that ℓ might behave like a white noise. Note that (3) does not imply exponential return of $Y - wL$ to zero, because its left-hand side differs from $d/dt(Y - wL)$ by the term $-\dot{w}L$. Sudden changes in w can change $Y - wL$, but sudden changes in L cannot.

It is probably easiest to accept this as a stylized representation of wage contracting if we restrict ourselves to thinking of cases with $\ell > 0$. Nonetheless, negative ℓ, which we must allow if we are contemplating possible white-noise-like behavior for ℓ, can be interpreted. It implies that contracts are being deliberately ended at a rate greater than the exogenous decay rate δ. For these terminations to be valued at a market flow rate w unrelated to the history of contracting implies a kind of "voluntary separation" interpretation. Workers who leave retain severance pay, retirement benefits, or the like that compensate them for the difference in value between their existing contract with the firm and what they can obtain in a new contract at the market rate. There are examples of layoffs that have this character ("employee buyouts", "early retirement incentives"), though of course not all of them do.

FOCs:

$$\partial C: \qquad D_1U(C,1-L) = \lambda \cdot (1-2\gamma V) \tag{8}$$

$$\partial L: \qquad D_2U(C,1-L) = -w\dot{\mu} - \dot{w}\mu + (\beta+\delta)\,\mu w \tag{9}$$

$$\partial Y: \qquad \frac{\lambda}{P} = -\dot{\mu} + (\beta+\delta)\,\mu \tag{10}$$

$$\partial B_C: \qquad -\frac{\dot{\lambda}}{P} = \left(r - \beta - \frac{\dot{P}}{P}\right)\frac{\lambda}{P} \tag{11}$$

$$\partial M: \qquad -\frac{\dot{\lambda}}{P} + \left(\beta + \frac{\dot{P}}{P}\right)\frac{\lambda}{P} = \gamma\frac{\lambda}{P}V^2 \tag{12}$$

IV.1.1. *Firms.*

$$\max_{\pi,\ell,L,Y,B_F} \int_0^\infty e^{-\beta t}\phi\left(\pi_t\right)dt \tag{13}$$

subject to

$$\zeta: \qquad \pi \leq f(L) - \frac{Y}{P} + \frac{\dot{B}_F}{P} - \frac{rB_F}{P} \tag{14}$$

$$\nu: \qquad \dot{Y} - w\dot{L} \geq -\delta\left(Y - wL\right) \tag{15}$$

$$B \geq 0\,. \tag{16}$$

Notice that the firms face the same contracting constraint in (15) that consumers do in (3).

FOCs:

$$\partial \pi: \qquad \phi' = \zeta \tag{17}$$

$$\partial L: \qquad \zeta f' = -\dot{\nu}w - \nu\dot{w} + (\beta+\delta)\nu w \tag{18}$$

$$\partial Y: \qquad \frac{\zeta}{P} = -\dot{\nu} + (\beta+\delta) \tag{19}$$

$$\partial B_F: \qquad -\frac{\dot{\zeta}}{P} + \left(\beta + \frac{\dot{P}}{P}\right)\frac{\zeta}{P} = r\frac{\zeta}{P} \tag{20}$$

IV.1.2. *Government.* The government budget constraint is

$$\dot{B}_C - \dot{B}_F + \dot{M} + \tau P = r\left(B_C - B_F\right) \ . \tag{21}$$

The government must set two dimensions of the behavior of $B_C - B_F$, τ, M, r, and P. Following what has been the convention in macroeconomics, we will at first assume that it fixes exogenously a time path for M and chooses a policy for the primary surplus τ that prevents real debt from exploding.

IV.1.3. *Model without stickiness.* The model without stickiness, to which we will compare this model with sticky wages, is nearly the same. We replace the Y terms in the consumer and firm budget constraints with WL, where W is the spot wage rate, and drop the equations defining the contract structure, (3) and (15), to arrive at, as FOC's for individuals

$$\partial C: \qquad D_1 U = (1 + 2\gamma V)\frac{\lambda}{P} \tag{22}$$

$$\partial L: \qquad D_2 U = \lambda \frac{W}{P} \tag{23}$$

and for firms

$$\partial L: \qquad \zeta f' = \zeta \frac{W}{P} \tag{24}$$

The other FOC's, other than those with respect to Y, which are dropped, are unchanged.

IV.1.4. *Equivalence of the Models with and without Stickiness.* The models with and without stickiness have the same solutions for any given time path of M. There are several ways to see why this is true. One is to note that equations (8)-(12) and (18)-(20) in the model with sticky wages play the same role, in terms of their implications once Lagrange multipliers are solved out, as the smaller set of corresponding equations (22)-(23), (18), and (20) in the model without stickiness, namely to produce the relationships

$$\frac{D_2 U \cdot (1 + \gamma V)}{D_1 U} = f'(L) \tag{25}$$

and

$$-\frac{\frac{d}{dt} D_1 U}{D_1 U} + \frac{\gamma \dot{V}}{1 + \gamma V} = r - \beta - \frac{\dot{P}}{P} \tag{26}$$

Of course while it is obvious that (25) and (26) are implied by the FOC's of the model without stickiness, that they follow from the model with stickiness requires some argument, which we will lay out below. For now, though, note that these two equations, together with (5) (the definition of V), the liquidity preference relation

$$r = \gamma V^2 \tag{27}$$

(implied by (11) and (12)) and the social resource constraint

$$C \cdot (1 + \gamma V) = f(L) \tag{28}$$

(which is implied by the private constraints (2) and (14) and the government constraint (21)) determine the five variables C, L, V, r, and P from a given policy-determined path for M. Again, further argument is required to show that there is a unique stable solution to the system, and we will lay out the argument out below. But accepting for now that there is a unique stable solution, we see that this system has the same form for the models with and without stickiness. In other words, the mapping from the time path of M to the time path of the five variables C, L, V, r, and P is the same in both models.

Note that the system (25)-(28) does not involve the level of P, only $\dot{P}/P$, and that M enters the system only as a ratio to P, in (5). Thus given any solution for a given path $\{M_t\}_{t=0}^{\infty}$, we know that a solution for the rescaled path $\{\kappa M_t\}_{t=0}^{\infty}$ is obtained by changing the original P path to κP while keeping r, C, V, and L unchanged. Money is therefore neutral. It is not super-neutral, because changing the rate of growth of P generally changes r, therefore (by (27)) V and therefore C and L. The important point is that both these statements are true, in exactly the same way, whether prices are "sticky" or not.

We still need to verify that (8)-(12) and (18)-(20) do reduce to (25)-(28). Since (11) and (20) have the same form, they imply that

$$\frac{\zeta_t}{\lambda_t} = \frac{\zeta_0}{\lambda_0}\,, \tag{29}$$

all t. From this point on we will assume that the solution involves stable time paths for λ and ζ. It is possible under mild regularity conditions to prove that these Lagrange multipliers have stable paths, but the argument becomes quite technical, and is thus omitted.[8] Equations (10) and (19) both have unique stable solutions, of the form

$$\mu(t) = \int_0^{\infty} e^{-(\beta+\delta)s} \frac{\lambda_{t+s}}{P_{t+s}} ds \tag{30}$$

[8] We are in effect verifying that if the model has a stable solution, then it has a stable solution that is the same for the sticky wage and non-sticky wage models. The argument we make explicitly here does not rule out the possibility that there might be solutions to the sticky wage model that involve unstable paths for the Lagrange multipliers and correspond to no equilibrium of the non-sticky price model. However, as noted in the text, it is possible to rule this out under mild regularity conditions.

$$\nu(t) = \int_0^\infty e^{-(\beta+\delta)s} \frac{\zeta_{t+s}}{P_{t+s}} ds \tag{31}$$

and because of (29) we know that this implies that

$$\nu \equiv \frac{\zeta_0}{\lambda_0} \mu \, . \tag{32}$$

But then from (8)-(10) we get

$$\frac{D_2 U \cdot (1 + 2\gamma V)}{D_1 U} = \frac{-\dot{\mu} + (\beta+\delta)\,\mu}{\lambda}$$

$$= -\frac{\dot{w}}{w}\frac{\mu}{\lambda} + \frac{-\dot{\mu} w + (\beta+\delta)\,\mu w}{\lambda} = -\frac{\dot{w}}{w}\frac{\mu}{\lambda} + \frac{w}{P} \tag{33}$$

while from (18)-(19) we get

$$f' = -\frac{\dot{w}}{w}\frac{\nu}{\zeta} + \frac{-\dot{\nu} w + (\beta+\delta)\,\nu w}{\zeta} = -\frac{\dot{w}}{w}\frac{\nu}{\zeta} + \frac{w}{P} \, . \tag{34}$$

From (29) and (32) we know that the right-hand sides of (33) and (34) are the same, so we arrive at (25), which was our target. It is easy to see that (26) is derived in exactly the same way in both models.

IV.1.5. *Discussion.* In this economy the wage on new contracts, w, is forward-looking. We can see from (34) that it can be represented as a weighted average of expected future marginal products (in dollars) of labor. At the same time, the aggregate economy-wide average wage, W, is by (6) and (7) a weighted average of current and past w's, which makes it a two-sided moving average of past and expected future marginal products (in dollars) of labor. This is exactly the form of wage dynamics postulated in the overlapping contract model of Taylor and the randomly-timed-adjustment model of Calvo. How is it that this model can mimic the price behavior of these other models, without producing their conclusion that stickiness generates non-neutrality?

In this economy, a labor contract promises a fixed stream of labor hours (subject to a randomly timed termination) in return for a fixed stream of dollar payments. The model contains a capital market that can produce present values for streams of future payments. The representative family, when deciding whether to increase or decrease L, considers by how much the present value of the stream of payments it obtains by increasing L by one unit exceeds the present value of the stream of future labor hours it is promising. This difference is the dollar return to decreasing leisure by one unit at this instant. It plays the same role as does the nominal wage in a model with spot labor markets.

Now suppose there is an unanticipated shift in M, resulting in a new equilibrium with a higher price level. Surely, it would seem, the outstanding stock of wage contracts, negotiated under different expectations, will produce real effects in the wake of this policy action. In a sense there are real effects. The representative family takes a capital loss on its old labor contracts, as these promise labor at a lower real rate of compensation than was originally anticipated. But this capital loss is the other side of a capital gain for the representative firm, which can now pay labor a lower than anticipated real rate of compensation. In this economy, the representative firm is owned by the representative family, which receives dividends from it. Thus these offsetting capital gains and losses have no effect on the representative family's budget constraint. And because the marginal compensation to changes in L is entirely a forward-looking object, the offsetting capital gains and losses produce no effect on labor supply or demand.

This result is a little like Ricardian Equivalence in that, on the one hand, it requires strong assumptions that are easy to criticize, while, on the other hand, it provides a limiting case that changes the way we think about macroeconomic issues. Few people think that Ricardian Equivalence holds even as a first-order approximation. Yet the principle that underlies it suggests that the effects of fiscal policy could well be much weaker than naive approaches that do not trace out complete general equilibrium effects would suggest. In this model of "neutral stickiness", perfect insurance and perfect capital markets are important maintained assumptions, as for Ricardian Equivalence. It is unlikely that the model is a good approximation to the truth. But by showing that the degree of stickiness of prices has no necessary connection to the amount of non-neutrality in the economy, it may change our thinking. With this model in mind, we see that no amount of micro-economic evidence for slowly or infrequently changed prices and wages can, by itself, tell us how important non-neutrality might be in the economy.

The conclusions of this model would be robust under a variety of changes in the nature of the contracts assumed. The essential feature of the contracts in this model, that leads to their not affecting neutrality, is that though they fix a "price" they do not represent open-ended commitments to sell at that price. Individual workers are working or not. If their contract specifies a wage that turns out to have been high, they benefit from this, and if their wage turns out to have been low, they lose. But these are changes in their wealth, not in effective slopes of their budget lines.

That there are no effects at all in this model depends on the assumption that there is a representative family that owns a representative firm, so that wealth redistribution between family and firm does not affect even the level of the budget line. More generally, with incomplete markets and nominal contracts of this type there would be wealth redistribution as the result of unanticipated price level changes, and this would have some real effect. The aggregate effects through this channel alone would probably be modest and of ambiguous sign, however.

Though wages are more often thought of as set by contracts something like those we describe here, many prices are also set along these lines. Catalogs, which offer goods at a price that is often fixed explicitly for a certain span of time, usually represent contracts of this type, because they usually do not represent a commitment to supply the goods promptly even if current stock runs out, and the goods being offered usually have some durability, so that both seller and buyer have some capacity to hold inventories. Thus if the market price moves much above that in the catalog, the stock of the good sells out, the good is unavailable from the catalog, and the effective shadow price is unaffected by the "fixed" catalog price. The rise in the market price has generated a wealth loss for the catalog issuer and a wealth gain for the consumers who snapped up the stock at the low catalog price. Modifying the model so that expenditures on goods, as well as income from wages, occur in fixed-price contracts that preserve neutrality, is straightforward.

V. A Menagerie of Stickiness Models

There are many ways to model nominal price stickiness. Some that seem very different in terms of the accompanying economic story-telling are not so different in their implications for aggregate behavior. Old fashioned and New Keynesian models form an example. Others that may seem similar in terms of the accompanying story-telling have different implications for aggregate behavior. The neutral stickiness model of section IV and Taylor-Calvo stickiness models form an example. In this section we describe in common notation an array of stickiness models. In the following section we display and compare the responses of these models to disturbances.

Each of the models splits the economy into a representative worker—or consumer, or household—and a representative firm. Each has the same time-separable utility function in consumption goods and leisure as in (1). Each has money demand generated by the appearance of real balances in the budget constraint of the consumer via a dependence of transactions costs per unit of consumption on velocity, as in the first term of (2). Each has firms maximizing the time-separable "utility" of dividends displayed in (13) and participating in the same bond market as consumers. Each has the same government budget constraint (21). In this section we show each agent's constraints and objectives. Display of the equations generated by FOC's from the optimization problems is left to appendix B.

V.1. **A More Realistic Classical Model with Stickiness.** Here we complicate the lean model of section IV above, by adding price stickiness and investment. There are adjustment costs for capital, so that there is a variable price of investment goods. In a model where transactions prices are not true shadow prices, there is some ambiguity as to how to define the transactions costs that determine the demand for money. It seems most natural to use a flow of actual contract payments to measure transactions, not a notional value of production or purchases priced at shadow prices. This eliminates the neat conclusion that contract prices have no effect on the real equilibrium.

Even without the complication of contract prices, a classical model with transactions costs implies some real effects of monetary policy. High inflation and nominal interest rates reduces equilibrium real balances and raises transactions costs. When the model includes an investment decision, it is possible in principle for investment decisions to be noticeably affected by variations in monetary policy even though transactions costs are not in equilibrium a large fraction of output (Rebelo and Xie 1997).

While it is therefore not a foregone conclusion that real effects of monetary policy will be small, in the model presented here, with realistic parameter values, the real effects of monetary policy do indeed emerge as small.

As with all the models we are considering, there is a representative consumer and a representative firm. Their objectives and constraints are described below.

V.1.1. *Consumer.*

$$\max_{C,L,V,M,B_C,Y_C,and Y_L} \int_0^\infty e^{-\beta t} \frac{C_t^{\mu_0} (1-L)^{\mu_1}}{\mu_0 + \mu_1} dt \tag{35}$$

subject to

$$\lambda: \quad Y_C + \dot{B}_C + \dot{M} + \tau \leq Y_L + \pi + rB_C \tag{36}$$

$$\psi_V: \quad V \geq \frac{Y_C}{M} \tag{37}$$

$$\psi_C: \quad C^* \geq C \cdot (1 + \gamma V) \tag{38}$$

$$\nu_{YC}: \quad \dot{Y}_C - p\dot{C}^* \geq -\eta_C \cdot (Y_C - pC^*) \tag{39}$$

$$\nu_{YL}: \quad \dot{Y}_L - w\dot{L} \leq -\eta_L \cdot (Y_L - wL) \tag{40}$$

$$B_C \geq 0, M \geq 0\,. \tag{41}$$

Two definitional equations used to generate transactions price and wage indexes, but not part of firm or consumer optimization problems, are

$$P \cdot C^* = Y_C \tag{42}$$

$$W \cdot L = Y_L\,. \tag{43}$$

V.1.2. *Firm.*

$$\max_{\pi, Y_C, C^*, Y_L, L, I, K, B_F} \int_0^\infty e^{-\beta t} \phi\left(\pi_t\right) dt \tag{44}$$

subject to

$$\zeta: \quad \pi \leq Y_C - Y_L + \dot{B}_F - rB_F \tag{45}$$

$$\omega: \quad C^* + \left(1 + \xi \frac{I}{K}\right) I \leq AK^\alpha L^{1-\alpha} \tag{46}$$

$$\sigma_K: \quad \dot{K} \leq I - \delta K \tag{47}$$

$$\sigma_{YL}: \quad \dot{Y}_L - w\dot{L} \geq -\eta_L \cdot \left(Y_L - wL\right) \tag{48}$$

$$\sigma_{YC}: \quad \dot{Y}_C - p\dot{C}^* \leq -\eta_C \cdot \left(Y_C - pC^*\right) \tag{49}$$

V.2. **Keynesian Stickiness: Optimizing Agents with Non-Market-Clearing Price Dynamics.** Here we take a classical model, like that of V.1 but without the contract prices and wages, and add to it a Phillips Curve and a Price Markup Equation. To make room for these two new equations, we drop the FOC's with respect to L of both worker and firm, implying that both take the evolution of employment as not directly under their control. The form of the Phillips Curve is then based on a measure of the gap in the L FOC of the worker, and the form of the Markup Equation is based on the gap in the L FOC of the firm. Nominal wages rise when workers are working more than they would like to at the given real wage, and prices rise when firms find their marginal labor costs exceed the value at current product prices of labor's marginal product.

A model like this does not have an elaborate story, based on forward-looking optimizing behavior, for its price and wage dynamics. New Keynesian style models do provide such stories. The stories are arguably in disagreement with microeconomic evidence, however. Also, if we see price sluggishness as ultimately arising out of limited information-processing capacity, dependence of model specification on elaborate modeling of forward-looking behavior in price setting may not be so attractive. Our objective here is to show that this somewhat old-fashioned and somewhat more transparent style of modeling can incorporate dynamically optimizing agents and in the end give results for macroeconomic behavior similar to those from New Keynesian models. Both types of model imply "jumpy" responses of real variables to nominal disturbances—in particular to monetary policy—that do not fit well with the facts.

V.2.1. *Consumer.*

$$\max_{C,L,V,M,B_C} \int_0^\infty e^{-\beta t} \frac{C_t^{\mu_0} (1 - L_t)^{\mu_1}}{\mu_0 + \mu_1} dt \tag{50}$$

subject to

$$\lambda: \qquad PC^* + \dot{B}_C + \dot{M} + \tau \le WL + \pi + rB_C \tag{51}$$

$$\psi_C: \qquad C^* \ge C \cdot (1 + \gamma V) \tag{52}$$

$$\psi_V: \qquad V \ge \frac{PC^*}{M} \; . \tag{53}$$

V.2.2. *Firm.*

$$\max_{\pi,I,K,L,C^*,B_F} \int_0^\infty e^{-\beta t} \phi(\pi_t)\, dt \tag{54}$$

subject to

$$\zeta: \qquad \pi \le PC^* - WL + \dot{B}_F - rB_F \tag{55}$$

$$\omega: \qquad C^* + \left(1 + \xi \frac{I}{K}\right) I \le AK^\alpha L^{1-\alpha} \tag{56}$$

$$\sigma_K: \qquad \dot{K} \le I - \delta K \; . \tag{57}$$

V.2.3. *Price and Wage Adjustment.*

$$\text{Phillips Curve:} \qquad \frac{\dot{W}}{W} = \eta_L \frac{\mu_1 U}{(1 - L) W \lambda} \tag{58}$$

Note that λ has the units of a real variable divided by P. See Appendix B.2.

$$\text{Markup Equation:} \qquad \frac{\dot{P}}{P} = -\eta_C \log\left(\frac{A \cdot (1 - \alpha)\, (K/L)^\alpha\, P}{W}\right) \tag{59}$$

V.3. **Monopolistic Competition with Menu Costs.** Aggregate behavior similar to that in the "old-fashioned" Keynesian model V.2 emerges from a model in which, rather than directly postulating wage and price adjustment equations, we allow firms to choose prices and workers to choose wages, but with quadratic penalties on the rates of change of prices and wages. As is standard in such models, we postulate monopolistic competition among price setters and (this is somewhat less standard) wage setters.

V.3.1. *Consumer.* The objective of the consumer is as in V.2, i.e. (50).

The consumer's budget constraint is

$$PC^* + \dot{B}_C + \dot{M} + \tau + \frac{1}{2}\chi\dot{W}^2 \leq WL + \pi + rB_C \tag{60}$$

Note that, though this constraint includes a penalty term on squared wage growth, and this generates terms of first-order importance in the FOC's, in the constraint itself, so long as we restrict attention to equilibria with constant steady-state prices, the wage growth penalty disappears in the linearization, as it contributes only a second-order term.[9]

Constraints (51)-(53) are unchanged. The new constraint

$$\psi_W: \qquad \frac{W}{\bar{\bar{W}}} = \left(\frac{L}{\bar{\bar{L}}}\right)^{-\eta_W} \tag{61}$$

is added.

V.3.2. *Firm.* The objective function is unchanged, given by (54). The cash flow and capital accumulation constraints (55) and (57) are unchanged, while the technical constraint (56) gains a price-change penalty term, to become

$$\omega: \qquad C^* + \left(1 + \xi\frac{I}{K}\right)I + \frac{1}{2}\chi\left(\frac{\dot{P}}{P}\right)^2 \leq AK^\alpha L^{1-\alpha} \tag{62}$$

The new constraint is:

$$\omega_P: \qquad \frac{P}{\bar{\bar{P}}} \leq \left(\frac{C^*}{\bar{\bar{C}}^*}\right)^{-\eta_P} \tag{63}$$

V.4. **Making Nearly Everything Sluggish: Adding Search and Limited Participation.** This model is an attempt to produce responses to disturbances that resemble what we see in the data, not only in sign but also in sluggishness of cross-variable responses. It has a variant of menu-cost price sluggishness as in the New Keynesian model V.3, but it adds additional inertia.

Neither labor input L nor consumption goods purchases C^* can be adjusted instantly in this model, by either firms or workers. Workers perceive themselves as able to influence the rates of growth of L and C^* by undertaking more or less search activity, x and z respectively. The results of search activity are assumed to dissipate,

[9] Since sustained non-zero inflation is commonly observed, one might think that it would be better to allow first-order effects into the budget constraint. However, it is a weakness of the menu-cost approach that it assumes that the costs of changing prices rapidly are "technologically" fixed. As a first approximation, it is probably more accurate to assume that it is costly to change prices at a rate different from the long run growth rate of prices than to assume that the form of the menu cost penalty is invariant to the steady-state inflation rate.

so x and z are non-zero in steady state. Firms perceive themselves as able to influence the rates of growth of L and C^* only by producing a price differential between their own firm's wage or price and the average wage or price. That is, they perceive the possibility of increasing their own rate of acquisition of new workers by raising their wage above the average wage or of increasing their own rate of consumption good sales by decreasing their price below the average price. Of course in the symmetric equilibrium we analyze, the firms in fact all choose the same values of W and P.

Note that this differs from the usual New Keynesian specification in that firms have no permanent market power. Demand for a firm's product declines indefinitely if it charges a price above the average level. The model has a kind of viscous perfect competition.

In addition, households perceive themselves as facing adjustment costs if they change their bond holdings. This is meant to capture the idea that money balances absorb short-term fluctuations in consumers' financial positions, with interest-bearing assets adjusting more smoothly. This has the important implication that, in contrast to all the other models we have discussed, in this one capital markets are incomplete. That is, the rate of return on bonds, which firms will equate to the rate of return on real investment, is, outside of steady state, not the same as the marginal rate of substitution between consumption at different dates faced by the consumer. In this respect the model is similar to the "limited participation" models (Christiano and Eichenbaum 1992, Fuerst 1992) elsewhere in the literature.

In keeping with our discussion of the implications of finite information-processing capacity, all this inertia is relative to levels that can themselves be subject to non-smooth shocks. This is meant to allow us to match the result from the data that there are non-smooth own shocks in most economic variables, with smooth cross-variable responses.

V.4.1. *Representative Household.*

$$\max \int_0^\infty e^{-\beta t} \frac{C^{\mu_0}(1-L^*)^{\mu_1}}{\mu_0+\mu_1}\,dt \tag{64}$$

with respect to C, C^*, L^*, L, B_C, M, V, z, x, Y_L, and Y_C, subject to

$$\lambda: \quad Y_C + \dot{B}_C + \dot{M} + \tau \le Y_L + \pi + rB_C \tag{65}$$

$$\psi_C: \quad C \cdot (1+\gamma V) \le C^* \tag{66}$$

$$\nu_{YC}: \quad PC^* \le Y_C \tag{67}$$

$$\nu_{YL}: \quad Y_L \le WL \tag{68}$$

$$\psi_L: \quad L + x\cdot(1+O_1 x) + z\cdot(1+O_2 z) + \tfrac{1}{2}O_3\dot{B}_C^2 \le L^* \tag{69}$$

$$\psi_V: \qquad \frac{Y_C}{M} \leq V \tag{70}$$

$$\psi_W: \qquad \frac{\dot{L}}{L} \leq G_1 \frac{x}{L} - \frac{G_2}{L} + \varepsilon_L \tag{71}$$

$$\psi_P: \qquad \frac{\dot{C}^*}{C^*} \leq H_1 \frac{z}{C^*} - \frac{H_2}{C^*} + \varepsilon_C \tag{72}$$

V.4.2. *Firm.*

$$\max \int_0^\infty \phi(\pi) e^{-\beta t}\, dt \tag{73}$$

with respect to W, P, L, K I, C^*, B_F, Y_L, Y_C, and π, subject to

$$\zeta: \qquad \pi \leq Y_C - Y_L + \dot{B}_F - rB_F \tag{74}$$

$$\omega: \qquad C^* + \left(1 + \xi \frac{I}{K}\right) I \leq \frac{AK^\alpha L^{1-\alpha}}{1 + \frac{1}{2} O_4 \cdot \left(\frac{\dot{W}}{W} - \varepsilon_W\right)^2 + \frac{1}{2} O_5 \cdot \left(\frac{\dot{P}}{P} - \varepsilon_P\right)^2} \tag{75}$$

$$\sigma_K: \qquad \dot{K} \leq I - \delta K \tag{76}$$

$$\sigma_{YL}: \qquad WL \leq Y_L \tag{77}$$

$$\sigma_{YC}: \qquad Y_C \leq PC^* \tag{78}$$

$$\omega_W: \qquad \frac{\dot{L}}{L} \leq G_1 \frac{x}{L} + \eta_L \log \frac{W}{\bar{W}} - \frac{G_2}{L} + \varepsilon_L \tag{79}$$

$$\omega_P: \qquad \frac{\dot{C}^*}{C^*} \leq G_3 \frac{z}{C^*} - \eta_C \log \frac{P}{\bar{P}} - \frac{G_4}{C^*} + \varepsilon_C \tag{80}$$

VI. Dynamic Behavior of the Models

We will consider first the response to a monetary expansion. In these calculations the fiscal policy rule is set at

$$\tau = -\phi_0 + \phi_1 \frac{B}{P}\,, \tag{81}$$

with $\phi_0 = .4$, $\phi_1 = .06$. Since $\beta = .05$, this is a "passive" or "Ricardian" rule. The monetary policy rule assumed for all the models is

$$\dot{r} = -\theta_2 \cdot (r - \beta) r + \theta_3 \log M + \varepsilon_M\,, \tag{82}$$

with $\theta_2 = .2$ in all cases. For models other than the "everything sticky" model V.4 we set $\theta_3 = .05$. However this policy is not feasible (at least with the fiscal rule we are using here) for model V.4, because it results in non-existence or non-uniqueness of equilibrium. (Sometimes one, sometimes the other, depending on other parameter settings.) In this case we set $\theta_3 = 0$.

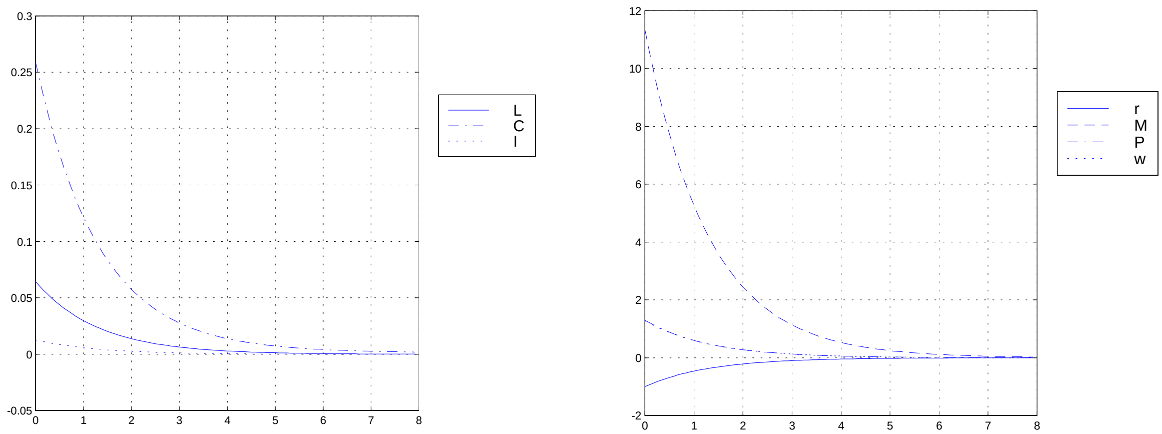

Figure 2. Classical Model: Responses to Monetary Policy Shock

We see in Figures 2 and 3 the responses of seven variables to an expansionary monetary policy shock in a purely classical model and in the classical model with neutral stickiness in prices and wages of section IV. As would be expected, the purely classical model has prices and wages responding proportionally to the monetary disturbance, and small responses of real variables. When we add neutral stickiness, the response of transactions prices of course is greatly reduced, but the real responses change only very slightly. The response in Figure 1 of industrial production to an interest rate disturbance shows a point estimate of about 0.4% for the maximal response to a 0.5% rise in the 3 month Treasury Bill rate. The Figure 1 results make no claim to reflecting a correct identification, but estimated real responses to identified monetary policy shocks are often close to estimated reduced form responses to interest rate innovations and are generally of this same order of magnitude. This estimated semi-elasticity of $.8 = .4/.5$ is about ten times the impact semi-elasticity of 0.06 for labor with respect to an interest rate change that appears in Figures 2 and 3. On the other hand, the empirical model's error bands include responses that are about as small as those in these classical theoretical models.

In Figures 4 and 5 we see the responses for two types of Keynesian model. Here the responses of real variables are large and instant, while the responses of nominal variables are smaller and delayed. The old-fashioned Keynesian model produces stronger and quicker responses, with some oscillation that is not present in the New Keynesian model's responses. However, there has been no attempt to calibrate the models so that they have similar behavior. They could be made to seem more similar by adjusting the parameters defining the degree of stickiness. Their qualitative behavior is similar: real variables jump strongly; nominal variables respond smoothly.

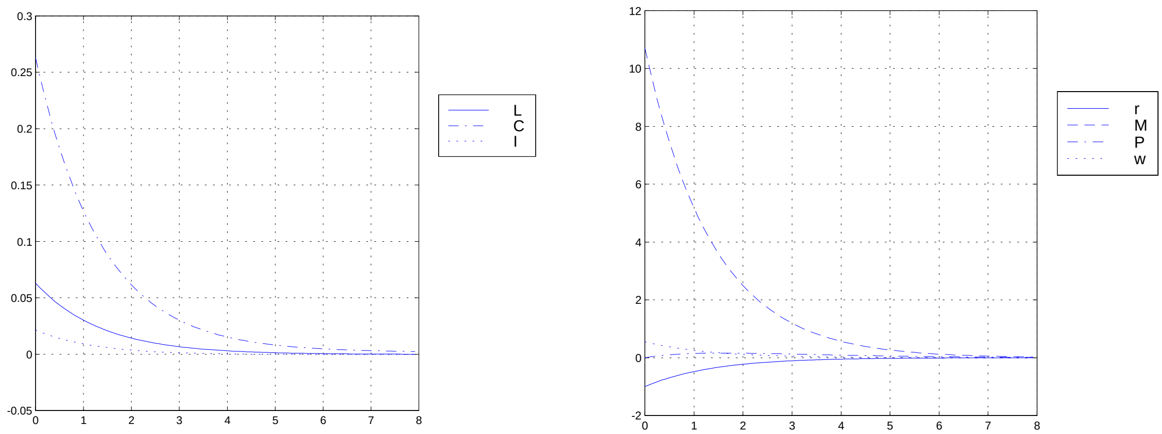

Figure 3. Sticky Classical Model: Responses to Monetary Policy Shock

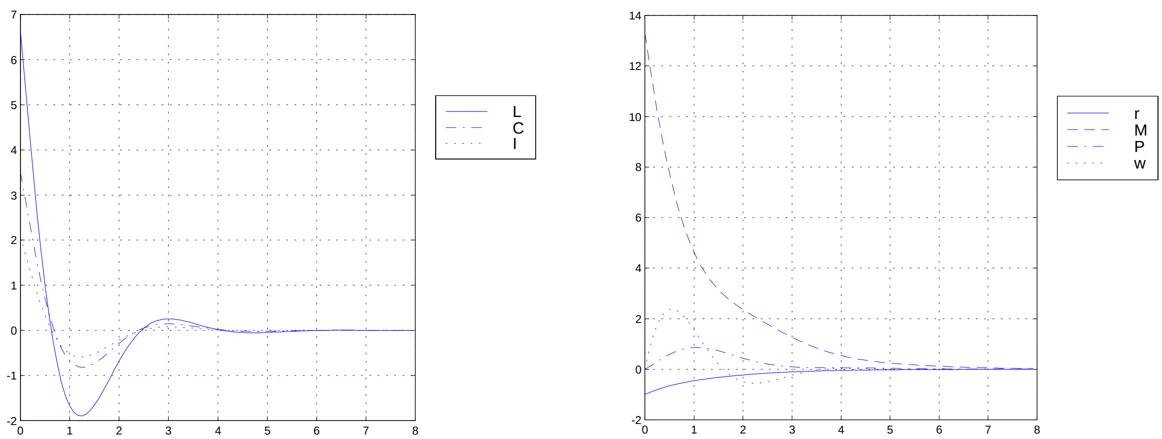

Figure 4. Keynesian Model: Responses to Monetary Policy Shock

In Figure 6 we see the responses for the "everything sluggish" model V.4. Here both real and nominal variables respond smoothly, and in contrast to all the other models M itself responds smoothly, as in estimated identified VAR's. The responses of the real variables and of M are all, as in the Keynesian models, larger than what we see in the data, but that could probably be fixed by estimation or calibration.

VII. Conclusion

The array of models examined here shows that shows that the qualitative behavior of an economic equilibrium model can match the observed behavior of prices, while delivering a wide range of implications for the effects of monetary policy on real variables. We have seen also that models with differing "microfoundations" for stickiness

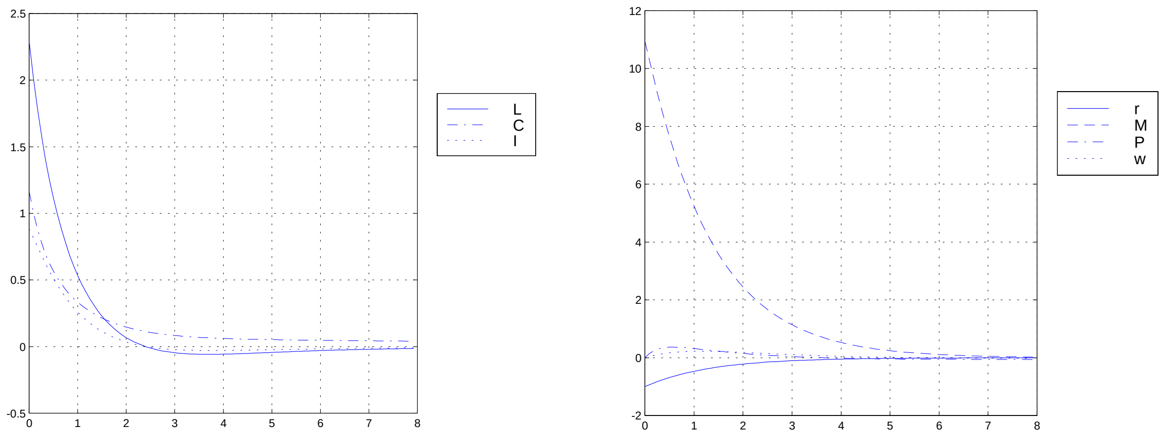

FIGURE 5. New Keynesian Model: Responses to Monetary Policy Shock

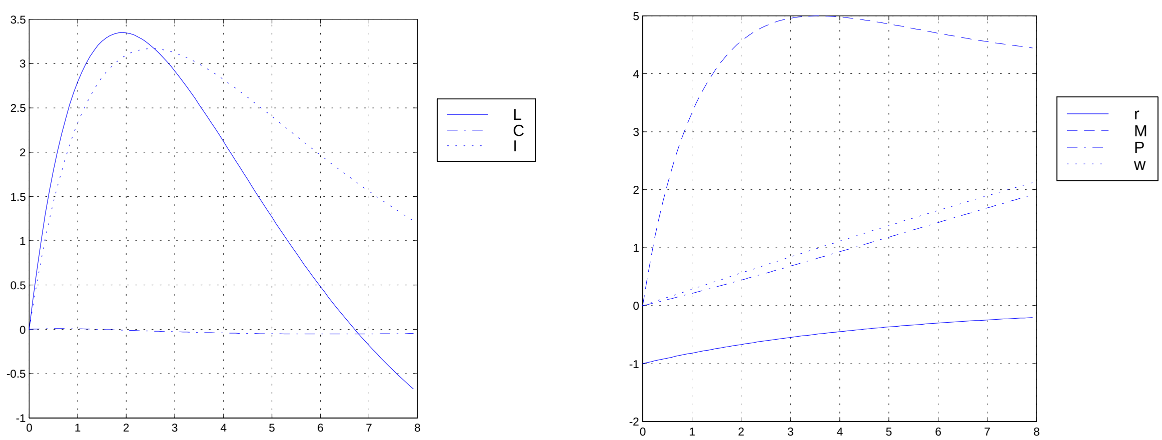

FIGURE 6. Everything Sluggish: Responses to Monetary Policy Shock

can be adjusted to produce similar implications for aggregate behavior. In this sense we are oversupplied with models of stickiness. Yet actual behavior of macroeconomic aggregates shows a combination of real and nominal sluggishness, together with quick response of nominal variables to own shocks, that does not fit easily into standard models of stickiness. Even qualitatively matching actual macroeconomic behavior seems to require a model with multiple sources of stickiness and delay. Such models are harder to solve and fit and less easily grasped by informal intuition than models that focus on single sources of friction or inertia. One conclusion from this paper's results, then, is that we should not let the convenience of simpler models as rhetorical devices outweigh their failures to match the data.

A second conclusion is that macroeconomists should rethink their commitment to modeling behavior as continuous dynamic optimization, with delays and inertia represented as emerging from adjustment costs. Delay and inertia seems to arise even where commensurate adjustment costs are difficult to see. Modeling economic agents as subject to constraints on information-processing rates may provide a route toward a more accurate theory of inertia without resort to ad hoc guesses about the nature of deviations from optimizing behavior.

Appendix A. A Brief Introduction to Information Theory

Here we present the basic ideas of information theory and some details of the argument that empirical observation of delayed, smooth reactions of variables in economic models to one another, combined with non-smooth responses of these variables to "own shocks", fits the hypothesis that individuals have limited information processing capacity. Under this hypothesis, the path connecting market signals to individuals' behavioral reactions should have the characteristics of a finite capacity *channel*, in the language of information theory.

A channel is something that takes as input a *signal* and produces an output. Once we have defined the possible input signals and described how they map into outputs (including the possibility that the mapping involves random noise) we can calculate what is called the channel's *capacity*. The central result of information theory is that it is possible to characterize the amount of information in any message we might want to transmit, and that if we allow a time for the transmission process equal to (amount of information)/(capacity), we can send the message with arbitrarily small error, even if the channel itself has substantial random noise in the connection of signal to output.

To make this concrete, we start with the simplest possible case. First we need to define the amount of information in a message. A message has to be characterized as the realization of a random variable. Or, in other words, we have to define the range of possible messages and their relative probabilities in order to define the amount of information in a message. In this simplest case, the message may be either 1 or 0, with probability $P[1] = p$. We characterize information by first characterizing the amount of lack of information, or *entropy* in a probability distribution. It turns out to be convenient to measure the entropy of a random variable z as $-E[\log_2(p(z))]$, where p is the p.d.f. of z. The use of base 2 logs is arbitrary, and the base measure against which the p.d.f. is constructed is also arbitrary, but otherwise it is possible to argue axiomatically that this is the only reasonable measure of entropy.[10] Thus

[10] If we change the base measure, the ranking of distributions by entropy will change. For example, our base measure, instead of putting equal discrete weight on 0 and 1, with no measure elsewhere,

for our two-point distribution, the entropy is $p\log_2(p) + (1-p)\log_2(1-p)$. This expression converges to zero as p approaches zero or one, and is maximized at $p = .5$, where it is just equal to one. The entropy of a two-point distribution with equal probabilities on the two points is called one *bit* of information.

Now we consider the simplest possible channel. Time t is discrete. The channel can take as its signal s at any date t only two possibilities: $s = 1$ or $s = 0$. The mapping from signal to output x is trivial: $s = 0 \Rightarrow x = 0$, $s = 1 \Rightarrow x = 1$. In other words there is no random error. If at each date we send a signal s drawn from a two-point distribution with equal probabilities on 0 and 1, then the receiver of the output each period has his distribution over s converted from one with entropy 1 bit (before he sees the signal) to one with zero entropy (after he sees the signal). Thus the channel transmits one bit per time period.

Of course we could also draw s from a distribution with $p \neq .5$, in which case less than one bit per time period would be transmitted. This suggests that we are wasting channel capacity in transmitting this way. But suppose we actually need to send a stream of zeros and ones in which, say, 90% of the stream is ones. How can we avoid wasting capacity? Break the stream of zeros and ones into groups of, say 4. Map these sequences into sequences of s for transmission as follows:

message	s sequence	n	p	$n \cdot p$
1111	1	1	0.66	0.66
0111	0000	4	0.073	0.292
1011	0001	4	0.073	0.292
1101	0010	4	0.073	0.292
1110	0011	4	0.073	0.292
0011	01000	5	0.0081	0.0405
0101	01001	5	0.0081	0.0405

(as is implicitly assumed when we conclude that with $p = .5$ the distribution has 1 bit of entropy), could put weight 2 on x=1 and weight 1 on x=0. Then the p.d.f. of a distribution with $P[1] = .5$ would take the value .25 at $x = 1$ and .5 at $x = 0$ and the entropy of the distribution would be $I(p) + p$, where $I(p)$ is the standard definition of entropy which is 1 at $p = .5$. With this definition of entropy maximum entropy occurs not at $p = .5$, but at $p = \frac{2}{3}$. Nonetheless, it can be verified that the average rate of information flow through a channel that transmits 0 or 1 without error, drawing from a distribution with $p = .5$, is still 1 bit per second with this modified definition of entropy. More generally the base measure has no effect on the average rate of information flow, no matter what the distribution of s and no matter what the distribution of $x|s$, and thus no effect on the definition of channel capacity.

0110	01010	5	0.0081	0.0405
1001	01011	5	0.0081	0.0405
1010	01100	5	0.0081	0.0405
1100	01101	5	0.0081	0.0405
1000	011100	6	0.0009	0.0054
0100	011101	6	0.0009	0.0054
0010	0111100	7	0.0009	0.0063
0001	0111101	7	0.0009	0.0063
0000	0111110	7	0.0001	0.0007
mean no. of s's				
per 4 message elements		2.0951		

It is easy to see that this scheme transmits the message without error, while at the same time requiring about half as much time to send a given message, on average, as the naive method of sending an s sequence that matches that in the message. The naive system would send $-.9\log_2(.9) - .1\log_2(.1) = .47$ bits per time period, whereas the scheme in the table sends almost twice that, or about .9 bits per time period. This is not quite full channel capacity, of course. To get closer to full capacity we would have to create a coding scheme for message sequences longer than 4.

Suppose the channel contains noise. For example, suppose sending $s = 1$ results in output $x = 1$ only with probability .75, while with probability .25 it results in $x = 0$. Suppose that the probability of "error" is the same when a 0 is sent. Then a channel user whose distribution on s was equal-probability on zero and one before seeing x, would after seeing x have a conditional probability distribution characterized by $P[s = x|x] = \frac{2}{3}$. The entropy of this distribution is .9183, so the information transmitted is 1 (the entropy of the distribution before seeing x) minus .9183, or .0817. This channel transmits .0817 bits per unit time. This means that if we are willing to transmit about 1/.0817, or about 13, s's for every 0 or 1 in a message sequence that is half ones and half zeros, we can transmit the message at that rate with arbitrarily small error. The type of code that does this is a bit more complicated than that shown above, so no example is provided here. Texts in information theory or computer science discuss such "error-correcting" codes.

In macroeconomic modeling, the messages agents send each other via market signals are generally most naturally thought of as real numbers, not zeros and ones. If s is drawn from a probability distribution with a density on the real line, and if the channel is the kind of simple error-free channel we discussed first above, with

$s = x$ with probability one, the channel is physically unrealizable, because it has infinite capacity. This may seem paradoxical, but a real number, transmitted exactly, amounts to transmitting an infinite sequence of 0's and 1's without error. Every infinite sequence of 0's and 1's can be thought of as the binary representation of a real number between 0 and 1, so we can send any message of 0's and 1's through this channel in one time period, no matter how long the message.

But if the channel contains noise, it can transmit real-valued s, producing real-valued x, while transmitting a finite amount of information per unit time. For example, suppose $s = x + \varepsilon$, where ε is i.i.d. $N(0,1)$ and our distribution over s before seeing x is $N(0,1)$. The distribution of s conditional on x is $N(.5s, .5)$. The entropy of a $N(0,\sigma^2)$ distribution is

$$\frac{\log_2 e}{2} + \log_2 \sigma + \frac{\log_2 \pi + 1}{2} \,.$$

Therefore the information transmitted by observation of x is $\log_2 \sqrt{1} - \log_2 \sqrt{.5} = .5$ bits.

Note that it cannot be that the channel is characterized by the ability to transmit an s drawn from an arbitrary distribution on the real line with an additive error, since if this were true the relative error could be made arbitrarily small by coding so that the s's actually sent are drawn from distributions that make s with high probability either very large or very small. The channel would have to require either a bound on the range of s, or an error specification that makes the distribution of the error depend on the distribution of s.[11] But our interest here is in characterizing the information transmission rate implicit in a given relation of inputs signals to outputs, not in an analysis of optimal coding for given physical channel characteristics. With proper coding, transmission of a $N(0,1)$ signal with an additive $N(0,1)$ error can be accomplished with a channel that sends one 0 or 1, without error, per unit time, and the transmission can be at a rate of two signals per unit time. This is just another way of stating the fact that the information transmission is at the rate of .5 bit per time period.

We conclude that finite-capacity transmission of real numbers must involve random error. Now we proceed to argue that transmission of information in continuous time must also in a certain sense involve delay.

Suppose we have a signal s that follows a standard Wiener process, and that the output of information transmission is $x = s + \varepsilon$, with ε also following a standard Wiener process, independent of s. This implies transmission of information at an infinite rate. To see why, consider the changes in s, x, and ε over small time intervals

[11] The error must be specified so that it remains possible to confuse two s values, no matter how far apart they are. For example making the standard deviation of ε proportional to $1 + s$ works.

$((j-1)/n, j/n$. We use the notation that, e.g., $\Delta s_j = s(j/n) - s((j-1)/n)$. The changes Δs_j and $\Delta\varepsilon_j$ will be i.i.d. across time, at any one date t being uncorrelated with each other, jointly normal, both distributed as $N(0, 1/n$. The change in x over the interval will therefore be distributed as $N(0, 2/n)$. But then if Δs_j is treated as the signal and Δx_j as the output, we are in the situation already discussed, where observation of the Gaussian output reduces the standard deviation of the distribution of the unknown Gaussian signal by a factor of 2, and a half bit of information is transmitted. This half bit per transmission is the same no matter how large n has been chosen. Thus in any finite interval, arbitrarily large amounts of information can be passed through this channel. This will be true in any setting where the error and the signal are independent stochastic processes driven by stochastic differential equations of the same order with Wiener process forcing terms.

A situation where, in continuous time, information is transmitted at a finite rate, can be constructed as follows. Let s be a cumulated Wiener process, i.e.

$$s(t) = \int_0^t W(v)\,dv\,, \tag{83}$$

and let ε be a Wiener process independent of s. Consider what happens when, as above, we break a finite interval, say (0,1), into n equal subintervals and examine the joint distribution of the sequences of s_j's and x_j's, where (for example) $s_j = s(j/n) - s((j-1)/n)$. We have to use the full joint distribution of the sequences, rather than examine the information transmitted at each j, one at time, because here the transmission error and the signal have different serial correlation properties, so no differencing or other filtering operation can produce i.i.d. signal-output pairs. The information about the sequence $\{s_j\}$ transmitted by observation of the sequence $\{x_j\}$ can be found by computing the entropy of the unconditional distribution of the full s sequence and comparing it to the entropy of the conditional distribution, given the x sequence. Both distributions will be multivariate normal, and it turns out that the information transmitted is

$$\tfrac{1}{2}\log_2\left(|\Sigma_s| - \left|\Sigma_{s|x}\right|\right)\,,$$

where Σ_s is the covariance matrix of the unconditional distribution of the s_j sequence and $\Sigma_{s|x}$ is its covariance matrix conditional on the x_j sequence. This can be calculated directly, and it can be shown to converge to a constant as $n \to \infty$. The convergence is rapid. Simply treating the signal as the single value $s(1)$ and the output as $x(1)$ transmits .415 bits, breaking the interval into two pieces ($n = 2$) transmits .613 bits, $n = 20$ transmits .721 bits and $n = 80$ transmits .727 bits. Evidently, with this smooth signal and less smooth error, there are rapidly diminishing returns to sampling x more frequently.

These two examples are representative. When s and ε are generated by nicely behaved stochastic differential equations driven by Wiener processes, then if ε has at least as many derivatives as s, observation of x can transmit unbounded amounts of information in finite time, while if ε has fewer derivatives than s, observation of x over a finite interval can transmit only finite amounts of information, no matter how frequently x is sampled.

Readers familiar with frequency-domain theory may find it useful to observe that if S_x is the spectral density of the signal and S_ε the spectral density of the noise, the rate of information transmission is

$$-\int_{-\infty}^{\infty} \log_2\left(\frac{S_\varepsilon}{S_\varepsilon + S_x}\right) dx \,, \tag{84}$$

from which the requirement that $S_x(\omega)/S_\varepsilon(\omega) \xrightarrow[\omega\to\infty]{} 0$ is apparent.

In general, because the noise must dominate the signal at high frequencies, the best current estimate of the signal will involve some smoothing of the channel output, and hence some delay relative to the signal.

Suppose there is a signal, say an asset price Q, that we wish to model as influencing behavior via a finite-capacity channel, with the behavior being represented by a choice variable Y. The implication of the analysis we have just been through is that we expect the relation between Y and Q to take the form (if we are modeling with Ito processes)

$$dY = a(Q)dt + dW(t) \tag{85}$$

with the W in (85) independent of the martingale driving Q, or (if we are working with general linear systems)

$$Y(t) = \int_0^{\infty} a(s)Q(t-s)ds + \varepsilon(t) = a * Q(t) + \varepsilon(t) \,, \tag{86}$$

where the $*$ denotes convolution and ε is a white noise independent of Q. If instead the variance of ε or W were zero, or if the level rather than the time derivative of Y depended on Q, the relationship would imply information flow at an infinite rate.

Appendix B. Details of the Models

B.1. **Realistic Classical Stickiness Model.** This is the model whose objective functions and constraints are laid out in section V.1.

Consumer FOC's:

$$\partial C: \quad \frac{\mu_0}{\mu_0+\mu_1} C^{\mu_0-1} (1-L)^{\mu_1} = \psi_C \cdot (1+\gamma V) \tag{87}$$

$$\partial C^*: \quad -\dot{p}\nu_{YC} - p\dot{\nu}_{YC} + (\beta+\eta_C)\, p\nu_{YC} = \psi_C \tag{88}$$

$$\partial Y_C: \quad \lambda = -\dot{\nu}_{YC} + (\beta+\eta_C) - \frac{\psi_V}{M} \tag{89}$$

$$\partial L: \quad \frac{\mu_1}{\mu_0+\mu_1} C^{\mu_0} (1-L)^{\mu_1-1} = -w\dot{\nu}_{YL} - \dot{w}\nu_{YL} + (\beta+\eta_L)\, w\nu_{YL} \tag{90}$$

$$\partial Y_L: \quad -\dot{\nu}_{YL} + (\beta+\eta_L)\,\nu_{YL} = \lambda \tag{91}$$

$$\partial B_C: \quad -\left(\frac{d}{dt} - \beta\right)\lambda = r \tag{92}$$

$$\partial M: \quad -\left(\frac{d}{dt} - \beta\right)\lambda = \frac{\psi_V}{M^2} \tag{93}$$

$$\partial V: \quad \psi_V = \gamma C \psi_C \tag{94}$$

Firm FOC's:

$$\partial \pi: \quad \zeta = \phi'(\pi) \tag{95}$$

$$\partial Y_C: \quad \zeta = -\dot{\sigma}_{YC} - (\beta+\eta_C)\,\sigma_{YC} \tag{96}$$

$$\partial C^*: \quad \omega = -p\dot{\sigma}_{YC} - \dot{p}\sigma_{YC} + (\beta+\eta_C)\,\sigma_{YC}p \tag{97}$$

$$\partial Y_L: \quad \zeta = -\dot{\sigma}_{YL} + (\beta+\eta_L)\,\sigma_{YL} \tag{98}$$

$$\partial L: \quad -w\dot{\sigma}_{YL} - \dot{w}\sigma_{YL} + (\beta+\eta_L)\, w\sigma_{YL} = \omega\,(1-\alpha)\, K^{\alpha} L^{-\alpha} \tag{99}$$

$$\partial I: \quad \sigma_K = \omega \cdot \left(1 + 2\xi \frac{I}{K}\right) \tag{100}$$

$$\partial K: \quad -\left(\frac{d}{dt} - \beta\right)\sigma_K + \delta\sigma_K = \omega \cdot \left(\alpha K^{\alpha-1} L^{1-\alpha} + \xi \frac{I^2}{K^2}\right) \tag{101}$$

$$\partial B_F: \quad -\left(\frac{d}{dt} - \beta\right)\zeta = \zeta \cdot (r-\beta) \tag{102}$$

B.2. **Keynesian Model.** This is the model whose objective functions and constraints are laid out in V.2

Consumer FOC's:

$$\partial C: \quad \frac{\mu_0 U}{C} = (1+\gamma V)\,\psi_C \tag{103}$$

where U is instantaneous utility, the undiscounted integrand in (50).

$$\partial L: \qquad \frac{\mu_1 U}{1-L} = W\lambda \tag{104}$$

(Equation above not used directly.)

$$\partial C^*: \qquad P\lambda + \frac{P\psi_V}{M} = \psi_C \tag{105}$$

$$\partial B_C: \qquad -\dot{\lambda} + \beta\lambda = r\lambda \tag{106}$$

$$\partial M: \qquad -\dot{\lambda} + \beta\lambda = \psi_V \frac{PC^*}{M^2} \tag{107}$$

$$\partial V: \qquad \gamma C\psi_C = \psi_V\ . \tag{108}$$

Firm FOC's:

$$\partial I: \qquad \sigma_K = \left(1 + 2\xi\frac{I}{K}\right)\omega \tag{109}$$

$$\partial K: \qquad -\dot{\sigma}_K + (\beta+\delta)\,\sigma_K = \omega \cdot \left(\alpha \cdot A \cdot \left(\frac{K}{L}\right)^{\alpha-1} + \xi\frac{I^2}{K^2}\right) \tag{110}$$

$$\partial L: \qquad \zeta W = \omega \cdot (1-\alpha) \cdot A \cdot (K/L)^{\alpha} \tag{111}$$

$$\tag{112}$$

Equation above not used directly.

$$\partial B_F: \qquad -\frac{\dot{\zeta}}{\zeta} = r - \beta \tag{113}$$

$$\partial C^*: \qquad P\zeta = \omega\ . \tag{114}$$

B.3. **New Keynesian Model.** This is the model whose objective functions and constraints are described in section V.3.

Consumer FOC's

Equations (103) and (105)-(107) are unchanged. Equation (104) is modified to the form

$$\partial L: \qquad \frac{\mu_1 U}{1-L} = W\lambda - \eta_W\frac{\psi_W}{L} \tag{115}$$

There is a new first order condition:

$$\partial W: \qquad -\chi\left(\ddot{W}\lambda + \dot{W}\dot{\lambda}\right) + \lambda L = \frac{\psi_W}{W} \tag{116}$$

Note that the term in $\dot{W}\dot{\lambda}$ contributes only second-order effects and therefore drops out of the linearization of this equation (though it contributes terms to the linearizations of FOC's).

Firm FOC's

Equations (109)-(113) remain unchanged. The FOC (114) gets an additional term, to become

$$\partial C^*: \qquad P\zeta = \omega + \frac{\omega_P \eta_P}{C^*} \tag{117}$$

The new FOC is

$$\partial P: \qquad -\frac{\ddot{P}}{P^2}\chi\omega - \frac{\dot{P}}{P}\chi\dot{\omega} + \left(\frac{\dot{P}}{P}\right)^2 \chi\omega = \frac{\omega_P}{P} - C^*\zeta \tag{118}$$

where only the first term on the left-hand side contributes to the linearization.

B.4. **Everything Sluggish.** *Consumer FOC's:*

$$\partial C^*: \qquad P\nu_{YC} - \frac{\left(\frac{d}{dt} - \beta\right)\psi_P}{C^*} = \psi_C - \frac{\psi_P \cdot (zH_1 - H_2)}{{C^*}^2} \tag{119}$$

$$\partial L: \qquad \psi_L - \frac{\left(\frac{d}{dt} - \beta\right)\psi_W}{L} = W\nu_{YL} - \frac{\psi_L(xG_1 - G_2)}{L^2} \tag{120}$$

$$\partial Y_C: \qquad \lambda = \nu_{YC} - \frac{\psi_V}{M} \tag{121}$$

$$\partial Y_L: \qquad \nu_{YL} = \lambda \tag{122}$$

$$\partial x: \qquad \psi_L \cdot (1 + 2O_1 x) = G_1 \frac{\psi_W}{L} \tag{123}$$

$$\partial z: \qquad \psi_L \cdot (1 + 2O_2 z) = H_1 \frac{\psi_P}{C^*} \tag{124}$$

$$\partial V: \qquad \gamma C \psi_C = \psi_V \tag{125}$$

$$\partial M: \qquad \lambda r = \psi_V \frac{Y_C}{M^2} \tag{126}$$

$$\partial B_C: \qquad -\left(\frac{d}{dt} - \beta\right)(\lambda + O_3 \dot{B}_C \psi_L) = r\lambda \tag{127}$$

$$\partial C: \qquad \psi_C \cdot (1 + \gamma V) = \frac{\mu_0}{\mu_0 + \mu_1} C^{\mu_0 - 1}(1 - L^*)^{\mu_1} \tag{128}$$

$$\partial L^*: \qquad \frac{\mu_1}{\mu_0 + \mu_1} C^{\mu_0}(1 - L^{\alpha})^{\mu_1 - 1} = \psi_L \tag{129}$$

Firm FOC's:

$$\partial\pi: \qquad \zeta = \phi'(\pi) \tag{130}$$

$$\partial Y_C: \qquad \sigma_{YC} = \zeta \tag{131}$$

$$\partial Y_L: \qquad \zeta = \sigma_{YL} \tag{132}$$

$$\partial B_F: \qquad -\left(\frac{d}{dt} - \beta\right)\zeta = r\zeta \tag{133}$$

$$\partial C^*: \qquad \omega - \frac{\left(\frac{d}{dt} - \beta\right)\omega_P}{C^*} = P\sigma_{YC} - \frac{\omega_P \cdot (H_1 z - H_2)}{{C^*}^2} \tag{134}$$

$$\partial I: \qquad \left(1 + 2\xi\frac{I}{K}\right)\omega = \sigma_K \tag{135}$$

$$\partial K: \qquad -\left(\frac{d}{dt} - \beta\right)\sigma_K + \delta\sigma_K - \omega\xi\frac{I^2}{K^2} = \frac{\omega\alpha A\left(\frac{K}{L}\right)^{\alpha-1}}{\mathcal{A}} \tag{136}$$

$$\partial L: \qquad \frac{\left(\frac{d}{dt} - \beta\right)\omega_W}{L} + W\sigma_{YL} + \frac{\omega_W \cdot (G_1 x - G_2)}{L^2} = \frac{\omega A \cdot (1-\alpha)\left(\frac{K}{L}\right)^{\alpha}}{\mathcal{A}} \tag{137}$$

$$\partial W: \qquad -\frac{\left(\frac{d}{dt} - \beta\right)}{W}\left(\frac{AK^\alpha L^{1-\alpha}\omega O_4 \cdot \left(\frac{\dot W}{W} - \varepsilon_W\right)}{\mathcal{A}^2}\right) + L\sigma_{YL} = \frac{\omega_W \eta_L}{W} \tag{138}$$

$$\partial P: \qquad -\frac{\left(\frac{d}{dt} - \beta\right)}{P}\left(\frac{AK^\alpha L^{1-\alpha}\omega O_5 \cdot \left(\frac{\dot P}{P} - \varepsilon_P\right)}{\mathcal{A}^2}\right) + \frac{\eta_C \omega_P}{P} = C^*\sigma_{YC} \tag{139}$$

In all the expressions above we are using the notation

$$\mathcal{A} = 1 + \tfrac{1}{2}O_4 \cdot \left(\frac{\dot W}{W} - \varepsilon_W\right)^2 + \tfrac{1}{2}O_5 \cdot \left(\frac{\dot P}{P} - \varepsilon_P\right)^2 . \tag{140}$$

Note that the $\mathcal{A}$ expression contributes nothing to the linearized form of the model.

References

Barro, R. J. (1979): "Second Thoughts on Keynesian Economics," *American Economic Review*, 69(2), 54–59, Papers and Proceedings of the Ninety-First Annual Meeting of the American Economic Association.

Christiano, L. J., and M. Eichenbaum (1992): "Liquidity Eects and the Monetary Transmission Mechanism," *American Economic Review*, 82(2), 346–353.

Fuerst, T. S. (1992): "Liquidity, Loanable Funds, and Real Activity," *Journal of Monetary Economics*, 29(1), 3–24.

Hall, R. E., and D. M. Lilien (1979): "Efficient Wage Bargains Under Uncertain Supply and Demand," *American Economic Review*, 69(5), 868–879.

Keating, J. W. (1997): "Is Sticky Price Adjustment Important for Output Fluctuations?," Discussion paper, University of Kansas, University of Kansas.

Leeper, E. M., and C. A. Sims (1994): "Toward a Modern Macro Model Usable for Policy Analysis," *NBER Macro Annual*, pp. 81–117.

Leeper, E. M., C. A. Sims, and T. Zha (1996): "What Does Monetary Policy Do?," *Brookings Papers on Economic Activity*, (2).

Phelps, E. S. (1968): "Money-Wage Dynamics and Labor-Market Equilibrium," *Journal of Political Economy*, 76(4), 678–711, Part2: Issues in Monetary Research, 1967.

Rebelo, S., and D. Xie (1997): "On the Optimality of Interest Rate Smoothing," Discussion paper, Kellog School of Management, Northwestern University.

Robert E. Lucas, J. (1973): "Some International Evidence on Output-Inflation Tradeoffs," *American Economic Review*, 63(3), 326–334.

Sims, C. A. (1994): "A Simple Model for Study of the Determination of the Price Level and the Interaction of Monetary and Fiscal Policy," *Economic Theory*, 4, 381–99.

——— (forthcoming): "The Role of Interest Rate Policy in the Generation and Propagation of Business Cycles: What Has Changed Since the 30s?," in *Proceedings of the 1998 Annual Research Conference*. Federal Reserve Bank of Boston.

Sims, C. A., and T. Zha (1998): "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*.