Comments on "A Behavioral New-Keynesian Model" by Xavier Gabaix

John H. Cochrane*

October 24, 2016

1. The basic point

Gabaix (2016b) posits an apparently small modification to standard new-Keynesian model. For purposes of discussion, I will strip it down even further to what I think is the central element:

$$x_t = ME_t x_{t+1} - \sigma(i_t - E_t \pi_{t+1}) \tag{1}$$

$$\pi_t = M^f \beta E_t \pi_{t+1} + \kappa x_t \tag{2}$$

$$i_t = \phi \pi_t + \hat{\imath}_t. \tag{3}$$

The new parameters $M, M^f \in [0,1]$ represent a degree of behavioralism. People don't pay full attention to their future prospects when thinking about consumption and inflation today. For example, with $M = 0.5$, news that consumption will be 10% larger

[*] Hoover Institution, Stanford University, and NBER. These are comments presented at the October 21 2016 NBER Economic Fluctuations and Growth program meeting, at the Federal Reserve Bank of Chicago. You can find the current version of this document and the computer programs at http://faculty.chicagobooth.edu/john.cochrane/

or smaller than trend next year induces people to change consumption by only 5% this year.

The effect of Gabaix' innovation is clearer to see if I turn the equations around as follows:

$$E_t x_{t+1} = \frac{1}{M}\left[x_t + \sigma(i_t - E_t\pi_{t+1})\right] \tag{4}$$

$$E_t \pi_{t+1} = \frac{1}{M^f}\frac{1}{\beta}\left[\pi_t - \kappa x_t\right] \tag{5}$$

You can see roughly here what I will make precise below: By making consumption and output today *less* sensitive to expectations $M < 1$, Gabaix makes the equilibrium paths of future consumption and inflation *more* sensitive to initial conditions. $M < 1$ induces *instability* – eigenvalues greater than one – in the model dynamics.

The standard new-Keynesian model has multiple stable "sunspot" equilibria. The standard way to fix this difficulty is to specify active ($\phi > 1$) monetary policy to destabilize the economy, leaving only one equilibrium or initial condition π_t that corresponds to non-explosive set of expectations. But that solution to the problem cannot apply at the zero bound, may not apply at other times, and suffers from theoretical problems.

So the central ingredient in this paper is that $M < 1$ can substitute for $\phi > 1$ to induce instability, and hence determinacy, in the standard new-Keynesian model. That is a profoundly important and relevant innovation.

(These comments draw on observations I have made in a string of related papers, Cochrane (2011), Cochrane (2014a), Cochrane (2014b), Cochrane (2016a) and most of all for the issues here, Cochrane (2016b). These papers also include citations to the vast literature. I collect all the algebra in an Appendix.)

2. More carefully

Equations (4)-(5) aren't really right to make this point as they leave $E_t\pi_{t+1}$ on the right hand side. The right set of equations pulls all the $t+1$ elements to the left hand side, writing the model in standard form as

$$E_t\begin{bmatrix} x_{t+1} \\ \pi_{t+1} \end{bmatrix} = \frac{1}{\beta^f M}\begin{bmatrix} \beta^f + \sigma\kappa & \sigma\left(\beta^f\phi - 1\right) \\ -\kappa M & M \end{bmatrix}\begin{bmatrix} x_t \\ \pi_t \end{bmatrix} + \frac{\sigma}{M}\begin{bmatrix} \hat{\imath}_t \\ 0 \end{bmatrix}$$

or,

$$E_t z_{t+1} = A z_t + v_t$$

In the conventional case, $M = 1$, with $\phi < 1$, this model has one eigenvalue greater than one and one eigenvalue less than one. The eigenvalue greater than one is solved forward, and can be used to uniquely determine x_t given π_t. But we are left with multiple stable paths for inflation $\{\pi_t\}$.

Stability by itself is not a problem. But the standard model only restricts expected future variables, not their outcomes, given variables π_t, x_t today. Any value of $\delta_{t+1} \equiv \pi_{t+1} - E_t\pi_{t+1}$ is possible. When the model has stable dynamics, such "sunspot" shocks will melt away due to the stability of the system, so there is no way to rule them out. Any value of π_t corresponds to a sequence of expectations $\{E_t\pi_{t+j}\}$ that converges to zero, so changing those expectations can change π_t arbitrarily. (In the conventional interpretation, causality runs from expectations of the future to outcomes today.)

The condition that both eigenvalues of A are greater than one (explosive) is

$$\phi + \frac{(1-M)\left(1-M^f\beta\right)}{\kappa\sigma} > 1. \tag{6}$$

In the conventional case, $M = 1$, active monetary policy $\phi > 1$ makes both eigenvalues greater than one. Now the system is unstable, and all but one value of π_t or δ_t corresponds to explosive expectations $E_t\pi_{t+j}$. Ruling out such explosions, we regain determinacy.

But here we see in this condition that a lower M and a lower M^f can compensate for $\phi < 1$, and even $\phi = 0$, and also deliver two explosive eigenvalues.

(I use "instability" in its traditional engineering sense, referring to A eigenvalues greater than one. "Instability" "volatility" and "determinacy" are not the same thing. Gabaix writes the model in form $Az_{t+1} = z_t +$ shocks, so his A is the inverse of mine, with opposite eigenvalues. He uses "stable" to mean "stationary.")

3. Why this paper is important

Figure 1 presents recent history in the US. Since 2009, the Federal funds rate has been stuck at zero. The easiest interpretation of this episode is that interest rates can no longer move one for one with inflation, so $\phi = 0$ and we have been forced to live with "passive" monetary policy. Japan has been there for 20 years. We hit the zero bound and...*nothing happened*. Inflation is stable, and if anything less volatile than before.

This long period of quiet inflation at the zero bound – and Japan's longer period – poses a deep challenge to monetary economics. Old-Keynesian models (including Milton Friedman's 1968 AEA address) predict inflation is *unstable* under a peg. They predict a deflation spiral, which did not happen.

Conventional new-Keynesian models predict that interest rate pegs are stable, a big change in the standard doctrine, and one validated by recent events. But the multiple stable equilibria mean that the standard model, and its standard interpretation, predict sunspot volatility, which also did not happen.

That conventional model also produces policy paradoxes at the zero bound. Forward

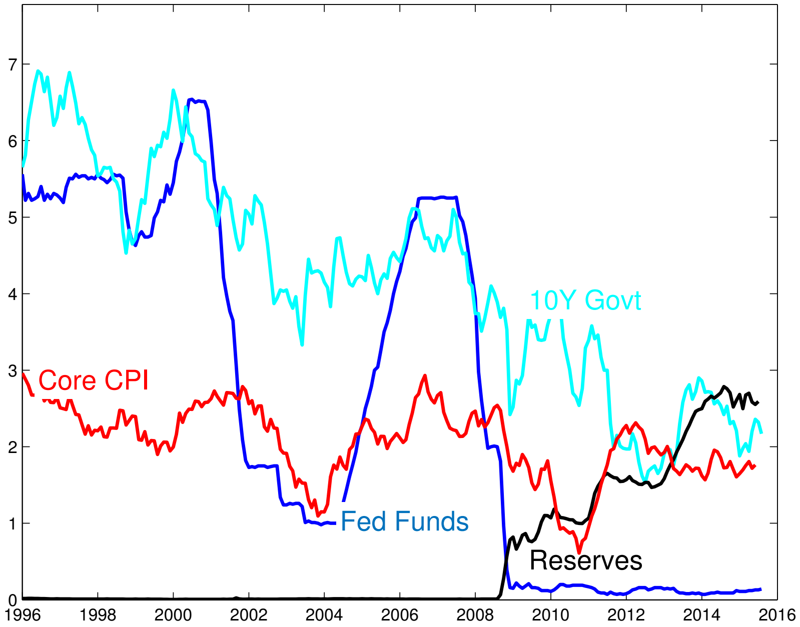


Figure 1: Recent US experience

stability means backward instability, so promises of policy changes further in the future have larger effects today. Less price stickiness speeds up dynamics, but that makes promises about the future more effective, taking the model further away from the frictionless outcome. Finally, this model predicts that higher interest rates *raise* inflation, both in the long run and the short run. This prediction may be correct, but at least the latter violates common belief, and one would like a model that is capable of delivering that belief. (I'll expand on all three issues in a minute.)

Gabaix offers a monetary theory that overcomes these deep problems. That's important!

While quiet inflation at the zero bound makes the problem pressing, policy has, in the standard new-Keynesian reading, been "passive" $\phi < 1$ many times in the past. New-Keynesian thought disparaged such episodes for inducing unnecessary sunspot volatility. But how much volatility? They really had little to say about inflation under passive policy. By restoring determinacy under passive policy, Gabaix offers one way to let the model talk productively about passive policy episodes.

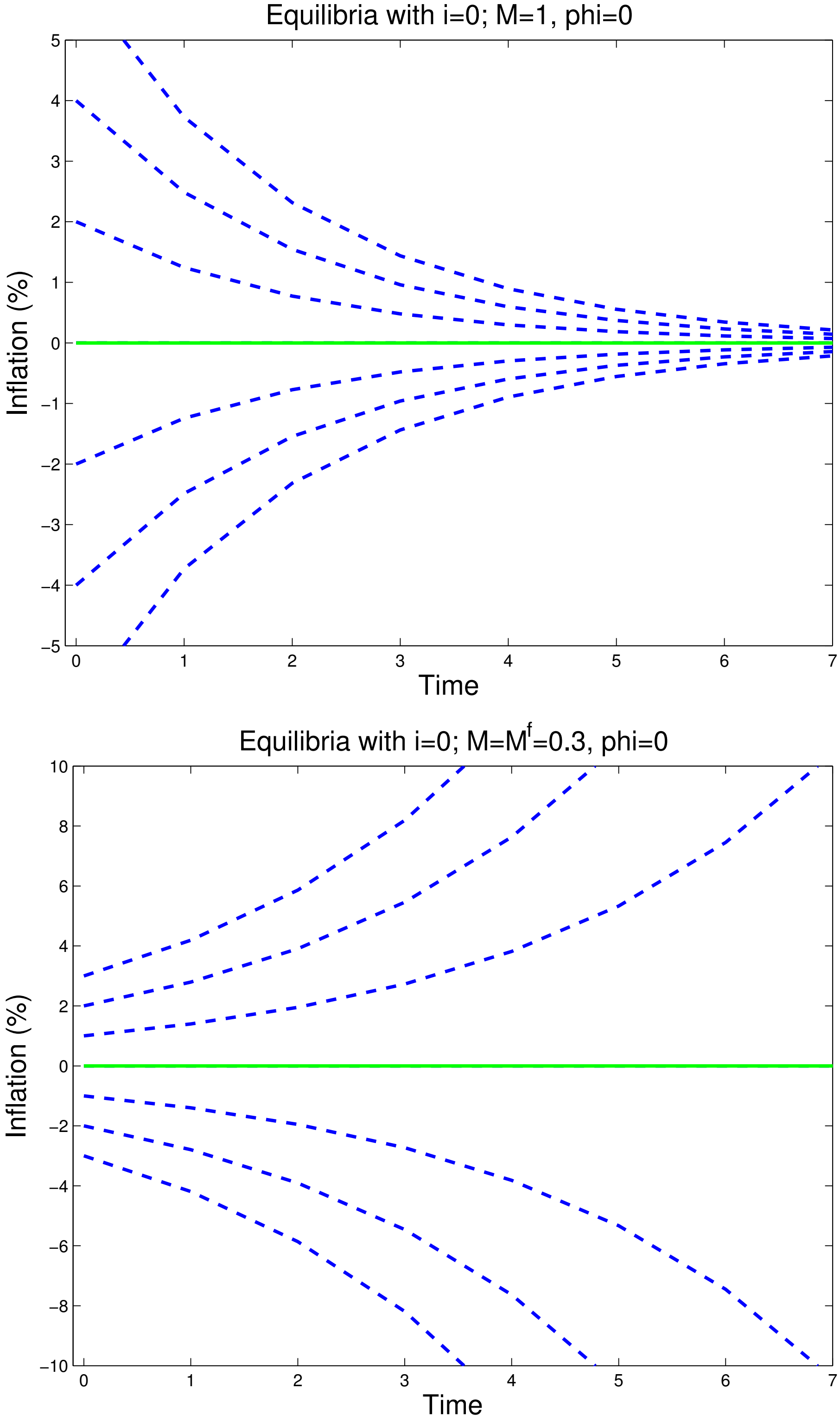
Furthermore, the $\phi > 1$ modification is not an uncontroversial answer to the stability and determinacy problem, even away from the zero bound. It requires that people believe the Fed deliberately destabilizes an otherwise stable economy to fight sunspots; these are never-seen and not-subgame-perfect off-equilibrum threats; and it requires a new rule against nominal explosions.

That theoretical debate may be less important now that we have spent 8 years at the zero bound, and Japan nearly 20. $\phi > 1$ is impossible at the zero bound, and the apparent stability of inflation at the zero bound is a decisive fact which models need to address. But Gabaix' modification applies always and everywhere, not just at the zero bound, so it offers a way to achieve determinacy that avoids the theoretical troubles of $\phi > 1$.

In sum, Gabaix' $M < 1$ takes the place of the hallowed Taylor principle $\phi > 1$ to deliver "good" properties of the new-Keynesian model, even at the zero bound, under an interest rate peg, or passve $\phi < 1$ monetary policy. It accounts for the low volatility and stability of inflation at the zero bound, avoids policy paradoxes, and also avoids the theoretical difficulties of $\phi > 1$. It's important!

3.1. Determinacy

To top plot of Figure 2 plots model solutions (deviations from the steady state) for a constant interest rate, $\hat{\imath} = 0$, in the standard case $M = 1$, and with an interest rate peg $\phi = 0$. Naturally, $x_t = 0$, $\pi_t = 0$ is a solution in this case. But there are multiple stable solutions, as shown. Furthermore, at each date, as well as date 0 as shown, the economy can unexpectedly jump from one of these equilibria to another via a sunspot shock δ_{t+1}.


Equilibria with i=0; M=1, phi=0

Equilibria with i=0; M=M^f=0.3, phi=0

Figure 2: Top: Stability and indeterminacy. Bottom: Gabaix fixes indeterminacy. Inflation under an interest rate peg (deviations from steady state). These are solutions of the standard new-Keynesian model for a constant interest rate, $\phi = 0$, deviations from steady state. Top: $M = 1$, bottom $M = M^f = 0.3$. $\kappa = \sigma = 0.5$; $\beta = 0.95$. The dashed lines represent multiple equilibria.

The bottom panel of Figure 2 shows the same situation with Gabaix' modification, $M = M^f = 0.3$ and still $\phi = 0$. The potential multiple equilibria of Figure 2 all explode going forward. Ruling out explosive equilibria, we are left with the unique equilibrium that just happens not to explode, $\pi_t = 0$, in this case. Dynamic *instability* induces *determinacy* (and hence stationarity, or "saddle path stability") in a model with expectational errors.

The conventional new-Keynesian answer induces explosive behavior with $\phi > 1$. Gabaix $M < 1$ can produce the same determinacy result, even when $\phi = 0$ at an interest rate peg.

3.2. Policy Paradoxes

Figure 3 illustrates the forward-guidance puzzle. Stability forwards means instability backwards. So, if policy can arrange a small change in expectations at some time in the future, that small change can have a very large change in outcomes today. (Typically, models specify that at a moment such as T=5 in the figure, the economy exits the zero bound and returns to active $\phi > 1$ policy. The active policy selects one equilibrium at $T = 5$, and working backwards that selects one equilibrium during the zero bound period. By changing the active policy – for example, the inflation target at $T = 5$ – such policy can thus change expectations of π_T and thus the outcome π_t.)

Furthermore, the *further* in the future the promised event, the *larger* its impact today. This behavior means that "forward guidance" announcements to manage expectations, if believed, can be very powerful. Fiscal stimulus, productivity or capital destruction – broken windows – can, by creating a little future inflation, create very large responses today.

If prices become less sticky, as κ rises, dynamics speed up. Faster forward stability is the same thing as faster backward explosions. So, *less* price stickiness makes all of these

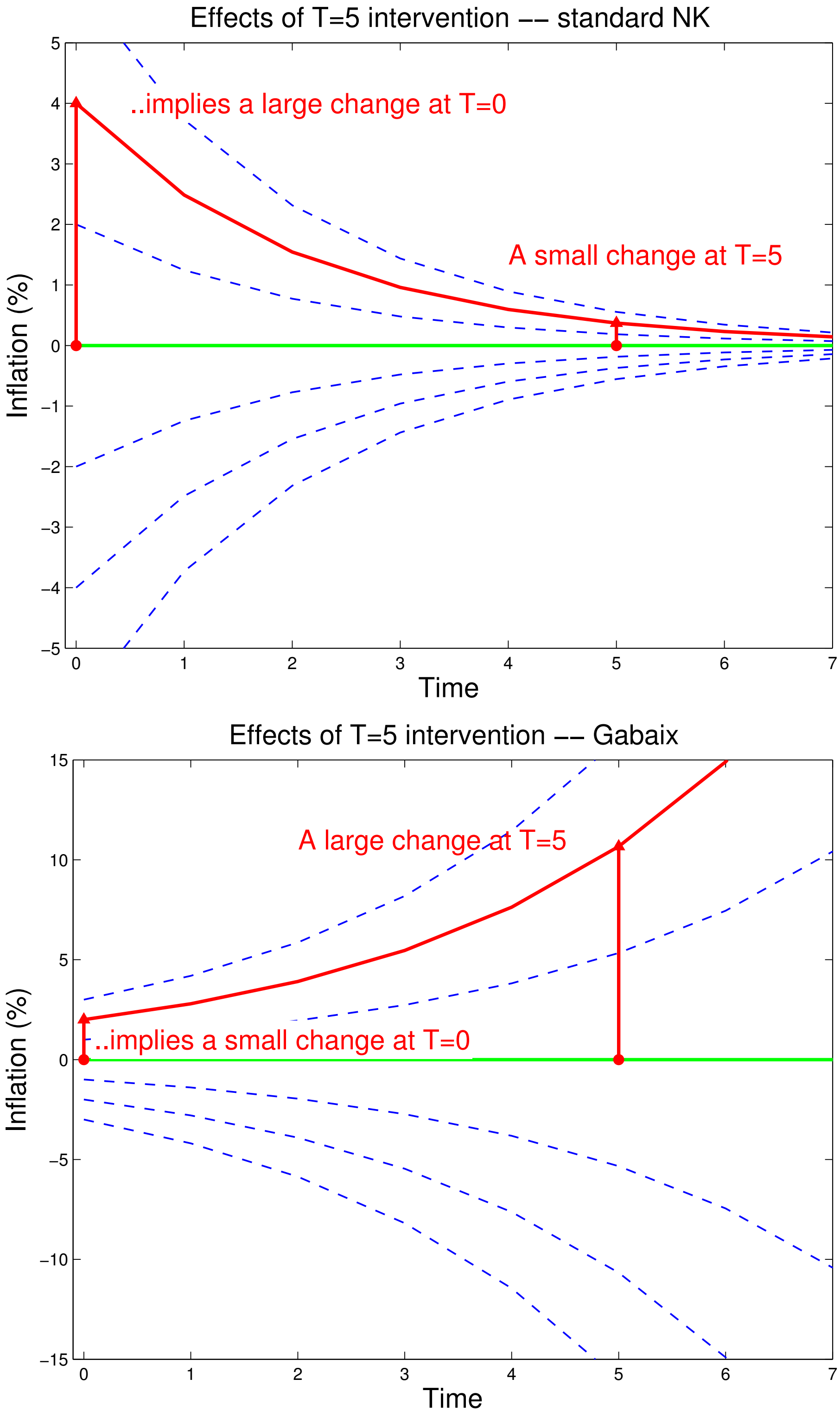


Figure 3: Top: Policy paradoxes. Bottom: Gabaix fixes the paradoxes. The solid line shows how a policy which changes expected inflation at $T = 5$ affects inflation at $T = 0$.

predictions *stronger*. (On these points, see Cochrane (2014b), Werning (2012).)

The bottom panel of figure 3 shows the same calculation with $M < 1$ and explosive eigenvalues. The forward-guidance paradox vanishes.

3.3. Fisherian responses

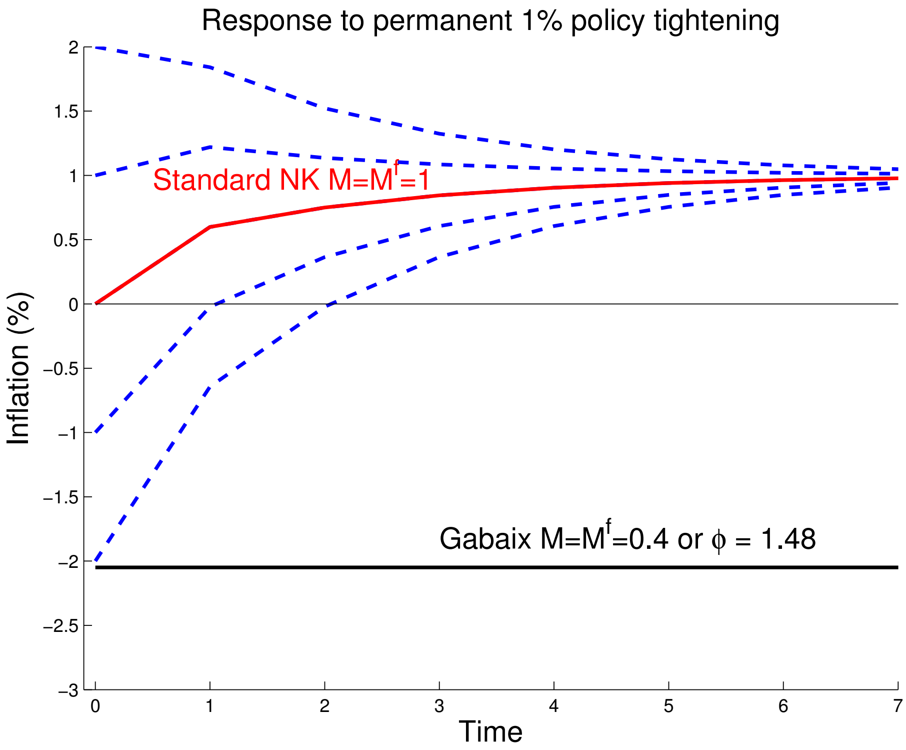


Figure 4: The Fisherian question. Inflation response to a 1% unexpected and permanent monetary policy disturbance î at time zero. Top lines: Standard new-Keynesian model with $M = 1$, $\phi = 0$. $\kappa = \sigma = 0.5$; $\beta = 0.95$. The solid line has no expectational jump $\delta_0 = 0$. The dashed lines add an expectational jump $\delta_0 = -2\%, -1\%, ... + 2\%$. Bottom line: Inflation response in a behavioral new-Keynesian model with $M = M^f = 0.4$. This is also the response function for the standard model with $M = M^f = 1$ and $\phi = 1.48$.

Finally, the standard model is both short- and long-run neo-Fisherian. Figure 4 shows the response of inflation to an unexpected permanent 1% monetary policy disturbance î). With no contemporaneous expectational jump $\delta_0 = 0$, a rise in interest rates produces uniformly higher inflation.

To get the conventional belief that higher interest rates at least temporarily lower inflation, one must somehow engineer an expectational jump $\delta_0 < 0$ coincident with the interest rate rise. But if you know how to induce expectational jumps, you can raise or lower inflation without bothering to change interest rates.

The bottom line of Figure 4 shows the response of inflation to the permanent 1% rise in the monetary policy shock $\hat{\imath}_t$, with Gabaix' modification $M = M^f = 0.4$. Now, the monetary policy tightening gives a gives a nearly 2% permanent reduction in inflation. The response of inflation to interest rates, formerly a two-sided moving average, now becomes entirely forward looking.

This response is identical to the response with $M = M^f = 1$ and $\phi = 1.48$. Here we see how Gabaix' innovation perfectly substitutes for active monetary policy to restore standard results even under an interest rate peg.

In fact, Gabaix' impulse-response improves on the standard $\phi > 1$ approach. If we achieve the plotted -2% inflation response to a monetary policy tightening $\hat{\imath}_t = 1\%$ via $\phi = 1.5$, the observed interest rate i_t *declines*. $i_t = \phi\pi_t + \hat{\imath}_t = 1.5 \times (-2\%) + 1\% = -0.5\%$. We observe interest rates *decline* by half a percent, inflation declines by 2%, and this counts as a "tightening" relative to the Taylor rule in which interest rates should have declined by 3%. By contrast, with $\phi = 0$, $M = 0.4$, the interest rate and the policy shock are the same, so observed interest rates rise by 1% to produce the inflation decline. (I've been writing "response to monetary policy shocks" rather than "response to interest rates" because of that little pathology of the standard model.)

3.4. Gabaix and old-Keynesian models

Gabaix' behavioral modification is *not* irrational or adaptive expectations, and this distinction is crucially important. Adaptive expectations models fail even worse than standard new-Keynesian models at the zero bound, and Gabaix' model solves this problem as well.

To illustrate in a compact way, I omit the intertemporal $E_t x_{t+1}$ IS term (adduce liquidity constraints, myopia, whatever). I substitute adaptive expectations $E_t \pi_{t+1} = \pi_{t-1}$ in the Fisher relation and Phillips curve, yielding a prototype "old-Keynesian" model of the type common in policy discussions:

$$\begin{aligned} x_t &= -\sigma(i_t - \pi_{t-1}) \\ \pi_t &= \pi_{t-1} + \kappa x_t \\ i_t &= \phi \pi_t + \hat{\imath}_t. \end{aligned}$$

The solution for inflation is

$$\pi_t = \frac{1+\sigma\kappa}{1+\sigma\kappa\phi}\pi_{t-1} - \frac{\sigma\kappa}{1+\sigma\kappa\phi}\hat{\imath}_t$$

With $\phi = 0$, this model is *unstable* but *determinate.* Since there is no expectational error, there is one equilibrium, and no variable can jump to a saddle path as consumption or asset prices do in forward-looking models.

In this model, the Taylor rule $\phi > 1$ produces a *stable* (coefficient on π_{t-1} is less than one) and *determinate* (one equilibrium) model. This is, I think, what the Fed and Taylor have in mind when they advocate such a rule.

But once again, $\phi > 1$ cannot apply at the zero bound. This model makes an unambiguously false prediction that deflation must spiral or vortex out of control at the zero bound. This prediction is even more dramatically false than the standard new-

Keynesian models' prediction that inflation must suffer extra sunspot volatility.

The new-Keynesian model solves this problem by making inflation stable, at the cost of adding multiple equilibria. The fiscal theory approach keeps stability, and picks one equilibrium. Gabaix makes inflation unstable, but changes the model to determine expectations only, so inflation can jump to the one saddle path that does not explode.

4. Doubts

I think I have made the case that this model is important, and deep. In fact, I'll argue here that it's too important to be right.

4.1. Magnitudes and limits

From condition (6), we need

$$(1-M)\left(1-M^{f}\beta\right)>\kappa\sigma. \tag{7}$$

to obtain determinacy (two explosive eigenvalues) at $\phi=0$, the zero bound,

Figure 5 shows the range of $\{M,\kappa\sigma\}$ values that satisfy this condition. The larger area assumes that firms and individuals have the same bias, $M=M^{f}$, while the smaller area assumes that firms are rational $M^{f}=1$, and this area changes M only.

The top horizontal line reminds us of the rationality parameter $M=M^{f}=1$. Higher κ means less price stickiness – as κ rises inflation can move more and more without changes in output. The arrow on the right points to the frictionless limit $\kappa=\infty$.

The figure emphasizes that the determinacy region is bounded away from rationality and the frictionless case.

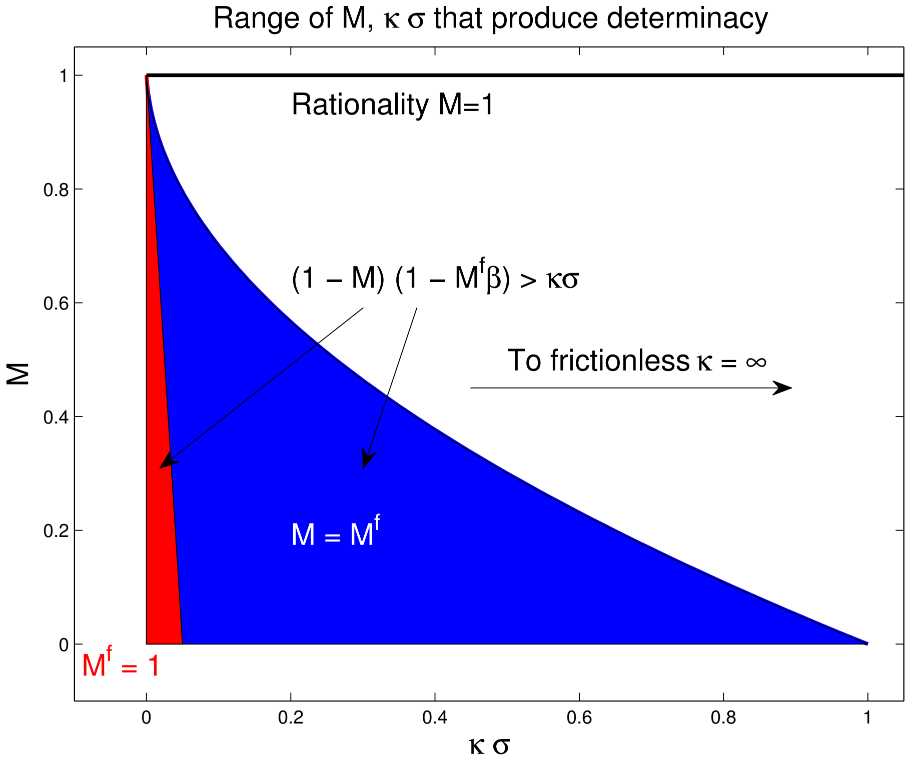


Figure 5: Parameter values that produce determinacy, $(1 - M)(1 - \beta M^f) > \kappa\sigma$, using $\beta = 0.95$. The larger region uses $M = M^f$, the smaller region uses $M^f = 1$.

A little bit of irrationality will not do. For any given $\kappa\sigma$, moving up and down in the graph, for the range near rationality $M = 1$, the eigenvalue remains stable and the model remains stable, indeterminate, neo-Fisherian, and puzzling. Or, starting in the determinacy region, as people become more rational, there always comes a moment at which the eigenvalue passes the boundary and we're back to stability, indeterminacy, and so forth.

A little bit of price stickiness will not do. For any given M, moving horizontally to the right in the graph, (6) as price stickiess is reduced sooner or later the eigenvalue passes the boundary and we're back to stability, indeterminacy, and so forth.

I find it a good principle that economic models should have smooth limits to the frictionless case. One could make a similar case that economic models should have smooth limits to rationality. This model has neither characteristic. Its basic properties rely on a discrete amount of irrationality and price stickiness.

In the simple model, you need a *lot* of behavioral bias, lots of price stickiness (low κ and little intertemporal substitution σ. Both κ and σ are on the order of magnitude of 1. κ measures how much inflation corresponds to a 1% change in output; and σ measures how much consumption growth response to a 1 percentage point change in interest rates. But with $\kappa = \sigma = 1$, even $M = M^f = 0$ is not sufficient to change the stability properties of the model. For illustrative calculations, I used $\kappa = \sigma = 0.5$. Even so, to produce the equivalent of the standard $\phi = 1.5$, I had to pick a severe amount of behavioral bias, $M = M^f = 0.4$. One needs especially low $\kappa\sigma$ to justify M anywhere near 1. Now, both κ and σ also can have downward behavioral bias too, so κ in the model is $m^f \kappa$ with $m^f \in [01]$. Such downweighting can expand the numerical range of true κ and σ. But it does not change the qualitative point that the region where the model works is bounded away from rationality or frictionless prices. And the point remains that one needs a lot of behavioral bias, either in a very low value of one M or in lots of different $M,\ M^f, m^f$.

But this quantitative point is less robust, and may change in more detailed versions of the model. The qualitative point is, I believe, fundamental.

4.2. Too important to be true

That Gabaix' model is bounded away from rationality has an easy and deep intuition. Gabaix' modification works by changing the basic stability properties of the model, changing eigenvalues less than one to eigenvalues greater than one, solving equations forward rather than backward. If you start with, say, an eigenvalue $\lambda = 0.7$, and you add some $M < 1$, you might raise the eigenvalue to $\lambda = 0.8$. But changing the eigenvalue a bit does *no good at all.* You have to keep adding irrationality, keep moving the eigenvalue, and only when you push the eigenvalue past 1.0 do you get *any effect at all* on the basic stability and determinacy properties that this model fixes.

This is both the models' deep strength, and its fundamental weakness. You have to swallow it whole or not at all. You can't take a little bite.

In the run of the mill new Keynesian model, we start with a model whose basic sign and stability properties are right, and add ingredients that help to address some other facts. For example, we might add time to build or habits to get a better shape of the impulse-response function. Little changes help a little. And one has the sense that the modifications are sufficient, but not necessary. We already have the basic signs and stability, and other modifications might produce the desired correlations.

Here, we study the minimal, necessary ingredients to understand basic sign and stability properties of monetary policy; the foundations on which minor modifications to match response functions will be built. Gabaix does not tell us that *small* behavioral changes to decision rules are *sufficient* to generate a model consistent with data, and classic views of the causal effects of monetary policy. Gabaix tells us that *large* behavioral distortions to decision rules are *necessary* to that project.

We are used to teaching undergraduates and offering policy analysis based on simplified stories for the effects of monetary policy, involving money supply and demand curves and rational agents. We like to think that these simple stories get the basic picture right, leaving magnitudes and dynamics for advanced elaboration.

If Gabaix is right, this presumption is utterly wrong. Large deviations from intertemporal optimization are at the heart of monetary policy. Monetary policy is not at heart supply-demand economics, like analyzing a price control or tax distortion, which might be quantitatively modified by a bit of irrationality. The basic *sign* and *stability* of monetary policy, including fundamental doctrines such as whether an interest rate peg is stable or determinate, whether interest rate increases raise or lower inflation, cannot be understood *at all* without irrationality. Trying to do otherwise, we are simply lying.

There are two kinds of economic policies: policies that rely fundamentally on basic supply and demand economics, but whose dynamics and real-world application may be influenced by behavioral and other frictions, and policies that fundamentally are based entirely on exploiting people's irrational behavior, using mechanisms that would be entirely absent in a supply-demand world. The analysis of taxes and tariffs are examples of the former. Keynesian multipliers are examples of the latter.

In Gabaix' world, monetary policy is *entirely* in the latter category. It is a magic trick, by which a super-rational benevolent policy-maker manipulates the behavioral misperceptions of us behavioral peasants. If we were only a bit more rational – responding to 80% of our expected future income, say, not 40%; or if the internet and procompetition reforms come along and lower price and wage stickiness in our economy, the basic *stability determinacy* and *sign* of monetary policy would suddenly change.

Behaviorists: Be careful what you wish for, you might get it! Are you really ready to go that far?

Similarly, If Gabaix' modification is right, it is always and everywhere right. Gabaix' modification is not a provision that one turns on for the zero bound and turns off otherwise, or invokes as an off-equilibrium threat to trim annoying multiple equilibria of theoretical models. If $M << 1$ and eigenvalues are unstable in 2016, then $M << 1$ and eigenvalues were unstable in 1976. He thus overturns not only zero-bound monetary economics but all monetary economics.

In particular, if Gabaix is right, then Clarida, Galí and Gertler (2000) are wrong. Their paper is a core – perhaps the core – empirical validation of the new-Keynesian model. They found that inflation stabilized along with a change from $\phi < 1$ in the 1970s to $\phi > 1$ in the 1980s. In their, standard, interpretation, this change in policy changed eigenvalues from stable to unstable, and the model from indeterminate, suffering sunspot volatility, to determinate and hence less volatile. If Gabaix is right, that interpretation of the 1980 stabilization is wrong – the eigenvalues were greater than one all along.

4.3. The foundations matter

Gabaix' form of behavioral bias is novel. The IS curve (1) is the more important ingredient,

$$x_t = M E_t x_{t+1} - \sigma(i_t - E_t \pi_{t+1}). \tag{8}$$

The central idea is that consumers react less than they should to state variables in their value functions, but then focus on one state variable – the deviation of income from trend, here – while being fully rational to others – changes in the trend, interest rates, probabilities, here. This is *not* a misperception of probabilities, or irrational expectations. It is *not* a mistake of excessive or hyperbolic discounting. It is not ambiguity, robust control, etc. It is not based on utility cost[1].

Getting to (8) is not easy. I've spent quite some time with Gabaix (2016b), Gabaix (2016a), and Gabaix (2014). If I had to pass a test asking me to "derive (8), " I would fail.

This is important. Suggestions like this one catch on if others can use them, and write more papers extending the idea. It is also vital to remove the taint of arbitrariness: Like Tolstoy's unhappy families, there are a thousand ways to be behavioral. Could someone else, making different assumptions about what is the "default" model, the agent's trend/cycle decomposition, specification of what is small vs. large error, and so forth, produce different distortions? Or, noting that gathering information to produce a rational expectation is expensive, might that person produce a more standard adaptive expectations models,

$$x_t = x_{t-1} - \sigma(i_t - \pi_{t-1})?$$

Would someone else, having read the first two Gabaix papers, and asked to produce "the" behavioral new-Keynesian model applying their principles, come up with the same

[1] As evidence that I am sympathetic to the general ideas, see Cochrane (1989).

thing?

Put a little less politely, if one dreamt up a convenient final form – for example, if one wanted to rescue the ISLM model one was taught at MIT in about 1978 – just how long would it take to dream up Gabaixian behavioral foundations to justify that desire? The key to any "theory" is not what it *can* produce, it is what it *cannot* produce, and that its predictions are inevitably *re*produced by fresh hands.

To be clear, I don't know the answers to these questions. I only know that I gave up in the time a even a masochistic discussant takes in thinking about a paper. My point is that these questions are important, and I hope followers will take them seriously.

Now, perhaps you will respond that it's the final model that matters, not the foundations. Much new-Keynesian modeling proceeds with abstract and unrealistic foundations. For example, the purely forward-looking Calvo Phillips curve is already microeconomically unrealistic, and empirically challenged. Theorists add "microfoundations" to add empirically useful lags of inflation. Empiricists put those in as well, but not really constrained by realistic theory or microfoundations. Perhaps all that matters in the end is linearized macro equations that work well. Perhaps we can consign three papers worth of foundations to the mental online appendix, and just proceed with $M = 0.4$.

I suspect in fact that's how this sequence of papers will be proceed. In lieu of Gabaix's detailed and thoughtful modeling of attention limits in an information-rich environment, sparse maximization, and so on, I forecast that his applied followers will simply say "suppose people pay less than full attention to expected future income in setting consumption today, so the IS curve is modified to (8). For details see Gabaix (2016b), Gabaix (2016a), Gabaix (2014), (etc.!) " and get on with the VAR and policy analysis – exactly as I have! "Add ME_tx_{t+1} with a small M to the IS equation? Sure," the poor souls operating FRBUS from some dungeon in the sub-basement of the Fed might answer – and not need 200 pages of equations to do it!

I think this outcome will be a shame.

Gabaix' modeling is much more than "suppose people pay less than full attention to expected future income." Most of all, it does not say that in all environments. Just which variables get the full rational treatment and which get discounted depends on the environment.

Moreover, if one is happy with relatively ad-hoc equations, we have plenty of them! One can just run a vector autoregression, bless the coefficients with behavioral or rule-of-thumb holy water to call them "structural," and proceed.

Or, just bestow some behavioral holy water on The old-Keynesian model, which remains the workhorse of policy analysis. This model is "behavioral" by substituting adaptive for rational expectations. (In the history of thought, I sense that the new-Keynesian model was greeted warmly because people thought, wrongly, that it simply supplied Lucas holy water to be sprinkled on the 1978 vintage ISLM model, justifying its continued use in the face of otherwise theoretical incoherence. Now that promise has been exploded, the demand for another font of holy water is strong.)

Foundations matter here, because the idea is so fundamental; necessary not just sufficient. If we were putting in habits to match a hump-shape of a response function, or adding indexing to put some lagged inflation in the Phillips curve (as Gabaix does later), it would not matter if the economics were a bit "as-if" or poorly microfounded. After all, we would only be polishing a model whose basic economic story, sign, and stability had exquisite micro-foundations.

But this modification is about getting the basic sign and stability properties of the model right; it is about the basic economic story of monetary policy. Casual, easily modified, what-if behavioralism does not belong in the basic story for sign and stability of monetary policy.

So, because the paper is so important, its foundations matter. I don't really understand them. But it is vital for anyone following in this track to really understand them, and not just add $M < 1$ as another behavioral what-if, with a forest of unread citations.

You can tell that this is a pretty hopeless quest. I didn't follow it. And Gabaix' own handling of the long run quickly gives up on careful behavioral foundations.

4.4. Long run

The stability of inflation, as summarized by Figure 1, and the convergence of inflation towards the nominal rate, seen especially in the Eurozone and Japan, suggests that inflation is Fisherian in the long run: higher interest rates must eventually lead to higher inflation. This is a natural version of long-run neutrality, which one might demand of a theory, or at least ask it to be able to express, as we would like a theory to be able to express the idea of a short-run negative sign.

Figure 6 presents again the responses to a permanent monetary policy shock, with some extra notation.

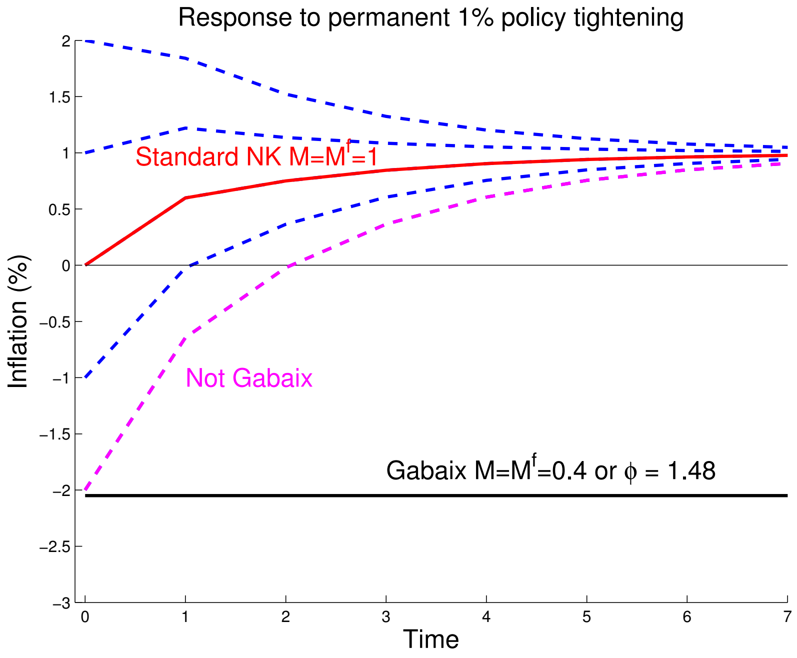


Figure 6: Response to a 1% permanent monetary policy shock.

In this context, we are reminded that Gabaix' model, like the standard $M = 1,\ \phi = 1.48$ model, produces a *permanent* decline in inflation in response to this shock.

One might have hoped that a model could select a different equilibrium, for example the highlighted "not Gabaix" equilibrium, of the original model. That one is really pretty – a temporary negative response, followed by a permanent positive Fisherian response.

But Gabaix' approach radically changes the stability properties of the model. It doesn't just help by picking a particular equilibrium of the original stable but indeterminate new-Keynesian model. It gets rid of the stability. The whole class of equilibria changes.

To restore long-run neutrality, Gabaix (section 5.3, and appendix section 9.2) "extends the model to have backward looking terms.." using a second behavioral distortion and a concept of "default inflation:" "In the microfoundation, each firm has two ways of predicting future inflation: one is via 'purely rational expectations', with π_t, another is via default inflation, π_t^d. The Phillips curve becomes

$$\pi_t = \beta M^f E_t \pi_{t+1} + \alpha \pi_t^d + \kappa x_t$$
$$\pi_{t+1}^d = \pi_t + \gamma(\pi_t^{CB} + (1-\zeta)\pi_t - \pi_t^d)$$

where π_t^d is "'default inflation' coming from indexation, and π_t^{CB}, the 'inflation guidance' by the central bank." Even so, Gabaix needs a delicate balance of parameters to get a long run Fisher relation.

To get the simple long-run neutrality result expected of the simplest rational and frictionless model, Gabaix has to add these epicycles of irrationality and price stickiness. This is now the *minimal* story one must tell undergraduates and policy-makers to explain why higher interest rates eventually must correspond to higher inflation.

Normally, we start with the Fisher relationship: $i_t = r_t + E_t\pi_{t+1}$. A frictionless model – and the frictionless limit point of the standard new-Keynesian model – is naturally

Fisherian in the short and long run. Higher interest rates lead to higher inflation, immediately and forever. One longs for a theory that naturally starts with this simple long-run neutrality in a frictionless and rational context, then adds frictions and details in a continuous way to produce the short-run negative sign and drawn out dynamics. A model that is local to frictionless, that has a smooth frictionless limit, can preserve this limit. But Gabaix' model takes a discrete jump away from frictionless and rational, so getting a positive long run response is much harder, requiring something like a second backwards jump. (The fiscal theory of the price level, adding long-term debt and price stickiness, does preserve stability, has a smooth frictionless limit, and allows one to select the pretty "not Gabaix" equilibrium.)

And Gabaix' approach to the long run violates the "take behavioral foundations seriously plea" I made in the last section, and appreciated of Gabaix' long string of papers. If we're allowed to add default inflation and central bank speeches so cavalierly, is adding $ME_t x_{t+1}$ really any better-founded? If willing to go here to get the long run, basic neutrality, is the alternative undisciplined behavioral fishing pond – say adding $\pi_t^e = (1-M)E_t\pi_{t+1} + M\pi_{t-1}$ and waving a wand of behavioralism over it – any better disparaged?

5. Globally necessary?

In sum, this paper is important! Standard models *utterly* fail to explain quiet inflation at zero interest rates.

Gabaix offers a fundamental change of the *basic story* of monetary economics. It is not a small modification on a basically correct model. A discrete and large amount of irrationality lies at the core of the basic sign, stability, and mechanism of monetary policy.

Specifically, a very wide class of standard new Keynesian models with $M = 1, \phi < 1$ are stable, indeterminate, rational, and produce sunspots and policy puzzles. For that wide class of models, Gabaix' $M << 1$ produces models that are unstable, locally determinate, and solve policy puzzles. His model is, however, bounded away from rational and frictionless, and needs epicycles for long run neutrality.

But are these assumptions really necessary? Is there no other model, outside the set considered by Gabaix, that can produce the desired behavior?

No. As I outline in Cochrane (2016b), adding (or, as I prefer to think about it, recognizing the presence of the) fiscal theory of the price level produces a model that is, even under an interest rate peg or zero bound, stable and determinate, produces the conventional belief about the short run response function (negative), and long-run Fisherian response (positive), is therefore consistent with our long period of quiet and stable inflation at near-zero interest rates, all while retaining rational expectations and intertemporal optimization. These properties of the model remain true as prices become less sticky, and in the frictionless limit. It is also considerably simpler than Gabaix' long-run model. I do not describe that model here – this is a comment on Gabaix, not the moment to introduce alternative models. The point here is that Gabaix' innovation – otherwise the only model left standing – is, though sufficient, not necessary. As Yoda might say, there is another model.

6. Appendix

6.1. Model in matrix form

Putting the model (1)-(3) in standard form, we have

$$E_t \begin{bmatrix} x_{t+1} \\ \pi_{t+1} \end{bmatrix} = \frac{1}{\beta^f M} \begin{bmatrix} \beta^f + \sigma\kappa & \sigma\left(\beta^f \phi - 1\right) \\ -\kappa M & M \end{bmatrix} \begin{bmatrix} x_t \\ \pi_t \end{bmatrix} + \frac{\sigma}{M} \begin{bmatrix} \hat{\imath}_t \\ 0 \end{bmatrix}$$

$$E_t z_{t+1} = A z_t + \frac{\sigma}{M} \begin{bmatrix} \hat{\imath}_t \\ 0 \end{bmatrix}$$

Here $\beta^f \equiv \beta M^f$. Gabaix writes this equation $z_t = AEz_{t+1}$+shock, so my A is the inverse of his.

6.2. Eigenvalues

One of the eigenvalues of A is always greater than one. Ruling out explosive solutions, there is therefore a linear combination $x_t = \alpha \pi_t$ that must always hold. The other eigenvalue, however, can be less than one, leading to multiple stable solutions. The eigenvalues of A are

$$\lambda = \frac{\left(M + \beta^f + \kappa\sigma\right) \pm \sqrt{\left(M + \beta^f + \kappa\sigma\right)^2 - 4M\beta^f\left(1 + \phi\kappa\sigma\right)}}{2M\beta^f} \tag{9}$$

The condition $\lambda > 1$ is therefore

$$\frac{\left(M + \beta^f + \kappa\sigma\right) - \sqrt{\left(M + \beta^f + \kappa\sigma\right)^2 - 4M\beta^f\left(1 + \phi\kappa\sigma\right)}}{2M\beta^f} = 1$$

which simplifies to

$$\phi + \frac{(1 - M)\left(1 - \beta^f\right)}{\kappa\sigma} > 1$$

6.3. Model Solution

While one can solve the model quickly via matrix techniques following the last subsection, here I use lag operator techniques to write the solution for inflation analytically. The more common matrix techniques scale better, but this approach allows one to see analytical solutions for simple systems.

The simple model is

$$\begin{aligned} x_t &= ME_t x_{t+1} - \sigma(i_t - E_t\pi_{t+1}) \\ \pi_t &= \beta^f E_t\pi_{t+1} + \kappa x_t \\ i_t &= \phi\pi_t + \hat{\imath}_t \end{aligned}$$

Substituting the Taylor rule,

$$\begin{aligned} x_t &= ME_t x_{t+1} - \sigma(\phi\pi_t + \hat{\imath}_t - E_t\pi_{t+1}) \\ \pi_t &= \beta^f E_t\pi_{t+1} + \kappa x_t \end{aligned}$$

Expressing the model in lag operator notation,

$$E_t(1 - ML^{-1})x_t = \sigma E_t\left(L^{-1} - \phi\right)\pi_t - \sigma\hat{\imath}_t$$

$$E_t(1 - \beta^f L^{-1})\pi_t = \kappa x_t$$

Forward-differencing the second equation,

$$E_t(1 - ML^{-1})(1 - \beta^f L^{-1})\pi_t = E_t(1 - ML^{-1})\kappa x_t$$

Then substituting into the first equation,

$$E_t(1 - ML^{-1})\left(1 - \beta^f L^{-1}\right)\pi_t = \kappa\sigma E_t\left(L^{-1} - \phi\right)\pi_t - \kappa\sigma\hat{\imath}_t$$

$$E_t\left[1 - \frac{M + \beta^f + \kappa\sigma}{1 + \kappa\sigma\phi}L^{-1} + \frac{M\beta^f}{1 + \kappa\sigma\phi}L^{-2}\right]\pi_t = -\frac{\kappa\sigma}{1 + \kappa\sigma\phi}\hat{\imath}_t.$$

We can stop here before solving to find the the long run impulse response function. With $L = 1$ we have

$$\left[1 - \frac{M + \beta^f + \kappa\sigma}{1 + \kappa\sigma\phi} + \frac{M\beta^f}{1 + \kappa\sigma\phi}\right]\pi_t = -\frac{\kappa\sigma}{1 + \kappa\sigma\phi}\hat{\imath}_t$$

$$\pi_t = -\frac{1}{\phi + \frac{(1-M)(1-\beta^f)}{\kappa\sigma} - 1}\hat{\imath}_t$$

A positive denominator is the condition that must be positive for both eigenvalues to be explosive. Therefore, the long-run response is negative. If the eigenvalues are close to one, the response function is very large. For $M = 1$:

$$\pi_t = -\left(\frac{1}{\phi - 1}\right)\hat{\imath}_t$$

Here we see the negative long-run impulse-response graphed in Figure 4.

Now, to solve the model. Factor the lag polynomial

$$E_t(1 - \lambda_1 L^{-1})(1 - \lambda_2 L^{-1})\pi_t = -\frac{\kappa\sigma}{1 + \kappa\sigma\phi}\hat{\imath}_t$$

where

$$\lambda = \frac{M + \beta^f + \kappa\sigma \pm \sqrt{(M + \beta^f + \kappa\sigma)^2 - 4M\beta^f\,(1 + \phi\kappa\sigma)}}{2\,(1 + \kappa\sigma\phi)}$$

These roots are the inverse of the eigenvalues given by 9. The system is stable and solved backward for $\lambda > 1$; it is unstable and solved forward for $\lambda < 1$.

6.4. Mixed case

When $\lambda_1 > 1$ and $\lambda_2 < 1$, reexpress the result as

$$E_t\left[(1 - \lambda_1^{-1} L)(1 - \lambda_2 L^{-1})\lambda_1 L^{-1}\right]\pi_t = \frac{\kappa\sigma}{1 + \kappa\sigma\phi}\hat{\imath}_t$$

$$E_t\left[(1 - \lambda_1^{-1} L)(1 - \lambda_2 L^{-1})\pi_{t+1}\right] = \lambda_1^{-1}\frac{\kappa\sigma}{1 + \kappa\sigma\phi}\hat{\imath}_t$$

The bounded solutions are

$$\pi_{t+1} = E_{t+1}\frac{\lambda_1^{-1}}{(1-\lambda_1^{-1}L)(1-\lambda_2 L^{-1})}\frac{\kappa\sigma}{1+\kappa\sigma\phi}\hat{\imath}_t + \frac{1}{(1-\lambda_1^{-1}L)}\delta_{t+1}$$

where δ_{t+1} is a sequence of unpredictable random variables, $E_t\delta_{t+1} = 0$ and $\delta_{t+1} = \pi_{t+1} - E_t\pi_{t+1}$.

Using a partial fractions decomposition to break up the right hand side,

$$\frac{\lambda_1^{-1}}{\left(1-\lambda_1^{-1}L\right)\left(1-\lambda_2 L^{-1}\right)} = \frac{1}{\lambda_1-\lambda_2}\left(1+\frac{\lambda_1^{-1}L}{1-\lambda_1^{-1}L}+\frac{\lambda_2 L^{-1}}{1-\lambda_2 L^{-1}}\right).$$

So,

$$\pi_{t+1} = \frac{1}{\lambda_1-\lambda_2}E_{t+1}\left(1+\frac{\lambda_1^{-1}L}{1-\lambda_1^{-1}L}+\frac{\lambda_2 L^{-1}}{1-\lambda_2 L^{-1}}\right)\frac{\kappa\sigma}{1+\kappa\sigma\phi}\hat{\imath}_t + \frac{1}{(1-\lambda_1^{-1}L)}\delta_{t+1}$$

or in sum notation,

$$\pi_{t+1} = \frac{\kappa\sigma}{1+\kappa\sigma\phi}\frac{1}{\lambda_1-\lambda_2}E_{t+1}\left(\hat{\imath}_t + \sum_{j=1}^{\infty}\lambda_1^{-j}\hat{\imath}_{t-j} + \sum_{j=1}^{\infty}\lambda_2^{j}E_{t+1}\hat{\imath}_{t+j}\right) + \sum_{j=0}^{\infty}\lambda_1^{-j}\delta_{t+1-j}.$$

For AR(1) shocks

$$\pi_{t+1} = \frac{\kappa\sigma}{1+\kappa\sigma\phi}\frac{1}{\lambda_1-\lambda_2}\left(\hat{\imath}_t + \sum_{j=1}^{\infty}\lambda_1^{-j}\hat{\imath}_{t-j} + \frac{\lambda_2}{1-\lambda_2\rho}\hat{\imath}_{t+1}\right) + \sum_{j=0}^{\infty}\lambda_1^{-j}\delta_{t+1-j}.$$

The impulse response function to a shock at time 0, announced at time 0, with $\hat{\imath}_t = 0$ $t < 0$; $\hat{\imath}_t = \hat{\imath}_0\rho^t$ for $t > 0$; and $\delta_t = 0$ for all t except $t = 0$, is then

$$\pi_{t+1} = \frac{\kappa\sigma}{1+\kappa\sigma\phi}\frac{1}{\lambda_1-\lambda_2}\left(\rho^t\hat{\imath}_0 + \sum_{j=1}^{t}\lambda_1^{-j}\rho^{t-j}\hat{\imath}_0 + \frac{\lambda_2}{1-\lambda_2\rho}\rho^{t+1}\hat{\imath}_0\right) + \lambda_1^{-(t+1)}\delta_0.$$

$$\pi_{t+1} = \frac{\kappa\sigma}{1+\kappa\sigma\phi}\frac{1}{\lambda_1-\lambda_2}\left(\rho^t + \rho^t\frac{\frac{1}{\lambda_1\rho}-\frac{1}{\lambda_1^{t+1}\rho^{t+1}}}{1-\frac{1}{\lambda_1\rho}} + \frac{\lambda_2}{1-\lambda_2\rho}\rho^{t+1}\right)\hat{\imath}_0 + \lambda_1^{-(t+1)}\delta_0.$$

$$\pi_{t+1} = \frac{\kappa\sigma}{1+\kappa\sigma\phi}\frac{1}{\lambda_1-\lambda_2}\left(\frac{1-\frac{1}{\lambda_1^t\rho^t}}{\lambda_1\rho-1} + \frac{1}{1-\lambda_2\rho}\right)\rho^t\hat{\imath}_0 + \lambda_1^{-(t+1)}\delta_0.$$

$$\pi_{t+1} = \frac{\kappa\sigma}{1+\kappa\sigma\phi}\frac{1}{\lambda_1-\lambda_2}\left(\left(\frac{1}{1-\lambda_2\rho}-\frac{1}{1-\lambda_1\rho}\right)\rho^t + \frac{1}{1-\lambda_1\rho}\lambda_1^{-t}\right)\hat{\imath}_0 + \lambda_1^{-(t+1)}\delta_0.$$

Note $\pi_0 = \delta_0$ only.

6.5. Both unstable case

When $\lambda_1 < 1$ and $\lambda_2 < 1$, write

$$E_t(1-\lambda_1 L^{-1})(1-\lambda_2 L^{-1})\pi_t = -\frac{\kappa\sigma}{1+\kappa\sigma\phi}\hat{\imath}_t$$

$$\pi_t = -E_t\frac{1}{(1-\lambda_1 L^{-1})(1-\lambda_2 L^{-1})}\frac{\kappa\sigma}{1+\kappa\sigma\phi}\hat{\imath}_t$$

$$\pi_t = E_t\frac{1}{\lambda_1-\lambda_2}\left(\frac{-\lambda_1}{1-\lambda_1 L^{-1}} + \frac{\lambda_2}{1-\lambda_2 L^{-1}}\right)\frac{\kappa\sigma}{1+\kappa\sigma\phi}\hat{\imath}_t$$

$$\pi_t = \frac{\kappa\sigma}{1+\kappa\sigma\phi}\frac{1}{\lambda_1-\lambda_2}E_t\left(-\lambda_1\sum_{j=0}^{\infty}\lambda_1^j\hat{\imath}_{t+j} + \lambda_2\sum_{j=0}^{\infty}\lambda_2^j\hat{\imath}_{t+j}\right)$$

$$\pi_t = \frac{\kappa\sigma}{1+\kappa\sigma\phi}\frac{1}{\lambda_1-\lambda_2}\left(-\lambda_1\sum_{j=0}^{\infty}\lambda_1^j\rho^j + \lambda_2\sum_{j=0}^{\infty}\lambda_2^j\rho^j\right)\hat{\imath}_t$$

$$\pi_t = \frac{\kappa\sigma}{1+\kappa\sigma\phi}\frac{1}{\lambda_1-\lambda_2}\left(-\frac{\lambda_1}{1-\lambda_1\rho} + \frac{\lambda_2}{1-\lambda_2\rho}\right)\hat{\imath}_t$$

$$\pi_t = \frac{\kappa\sigma}{1+\kappa\sigma\phi}\frac{1}{\lambda_1-\lambda_2}\left(\frac{\lambda_2\left(1-\lambda_1\rho\right)-\lambda_1\left(1-\lambda_2\rho\right)}{\left(1-\lambda_1\rho\right)\left(1-\lambda_2\rho\right)}\right)\hat{\imath}_t$$

$$\pi_t = -\frac{\kappa\sigma}{1+\kappa\sigma\phi}\left(\frac{1}{\left(1-\lambda_1\rho\right)\left(1-\lambda_2\rho\right)}\right)\hat{\imath}_t$$

Multiple solutions at an interest rate peg

I'll use perfect foresight starting at time 0. with $\hat{\imath} = 0$,

$$(1 - \lambda_{-} L^{-1})(1 - \lambda_{+} L^{-1})\pi_t = 0$$

we always have $\lambda_{-} < 1$, so I always solve that forward. Then the family of solutions is

$$(1 - \lambda_{+} L^{-1})\pi_t = 0$$

$$\lambda_{+}\pi_{t+1} = \pi_t$$

$$\pi_{t+1} = \lambda_{+}^{-1}\pi_t$$

These generate the multiple equilibria of Figure 2.

References

Clarida, Richard, Jordi Galí, and Mark Gertler, "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory,," *Quarterly Journal of Economics*, 2000, *115*, 147–180.

Cochrane, John H., "The Sensitivity of Tests of the Intertemporal Allocation of Consumption to Near-Rational Alternatives," *American Economic Review*, 1989, pp. 319–337.

_ , "Determinacy and Identification with Taylor Rules," *Journal of Political Economy*, 2011, *119* (565-615).

_ , "Monetary Policy with Interest on Reserves," *Journal of Economic Dynamics and Control*, 2014, *49*, 74–108.

_ , "The New-Keynesian Liquidity Trap," 2014. Manuscript.

_ , "Do Higher Interest Rates Raise or Lower Inflation?," *Manuscript*, 2016.

_ , "Michelson-Morley, Occam and Fisher: The Radical Implications of Stable Inflation at Near-Zero Interest Rates," *Manuscript*, 2016.

Gabaix, Xavier, “A Sparsity-Based Model of Bounded Rationality,” *Quarterly Journal of Economics*, 2014, *129*, 1661–1710.

__ , “Behavioral Macroeconomics Via Sparse Dynamic Programming,” *Manuscript*, 2016.

__ , “A Behavioral New Keynesian Model,” *Manuscript*, 2016.

Werning, Iván, “Managing a Liquidity Trap: Monetary and Fiscal Policy,” *Unpublished*, April 2012. Manuscript, MIT.