Journal of Monetary Economics 56 (2009) 1109–1113

Contents lists available at ScienceDirect

Journal of Monetary Economics

journal homepage: www.elsevier.com/locate/jme

Can learnability save new-Keynesian models?

John H. Cochrane [a,b,*]

[a] University of Chicago Booth School of Business, 5807 S. Woodlawn, Chicago, IL 60637, USA
[b] National Bureau of Economic Research, USA

ARTICLE INFO

Article history:
Received 21 October 2009
Received in revised form
26 October 2009
Accepted 27 October 2009
Available online 4 November 2009

JEL classification:
E30
E31
E42

Keywords:
New-Keynesian model
Learning
Learnability
Multiple equilibria
Indeterminacy

ABSTRACT

McCallum (2009) argues that "learnability" can save new-Keynesian models from indeterminacies. He claims the unique bounded equilibrium is learnable, and the explosive equilibria are not. However, he assumes that agents can directly observe the monetary policy shock. Reversing this assumption, I find the opposite: the bounded equilibrium is not learnable and the unbounded equilibria are learnable. More generally, I argue that a threat by the Fed to move to an "unlearnable" equilibrium for all but one value of inflation is a poor foundation for choosing the bounded equilibrium of a new-Keynesian model.

1. Introduction

McCallum (2009), extending a long literature and Evans and Honkapohja (2001) in particular, argues that "learnability" can save new-Keynesian models from their indeterminacies.

McCallum and I agree on the most important points. Standard new-Keynesian Taylor-rule models have multiple, economically valid solutions. There is no reason in standard economics to pick only the unique locally-bounded solution. If you want to throw out the other solutions, you need to add some new principle. "Learnability" is, potentially, an extra principle that could do the trick. I also applaud McCallum's effort to apply and interpret the rather abstract learnability literature in the specific context of these important models.

However, I do not think "learnability" does, in fact, solve the problem, at least as expressed in this paper. First, I think McCallum went wrong in applying Evans and Honkapohja's (2001) results. He assumed that the monetary policy shock is *directly* observable. By "directly," I mean that agents cannot try to recover it as a regression residual. Removing that assumption reverses the result: Under the Taylor principle, the unique locally-bounded equilibrium is, in fact, the only one that is *not* learnable. The explosive equilibria *are* learnable.

* Correspondence address: University of Chicago Booth School of Business, 5807 S. Woodlawn, Chicago, IL 60637, USA. Tel.: +1 773 702 3059; fax: +1 773 702 0458.
E-mail address: john.cochrane@chicagobooth.edu
URL: http://faculty.chicagobooth.edu/john.cochrane/research/Papers/

0304-3932/$ - see front matter
doi:10.1016/j.jmoneco.2009.10.012

Second, even if the learnability results hold, I still do not think the result provides a satisfactory model of inflation determination. Is inflation really determined at a given value because for any other value the Fed threatens to take us to a valid but "unlearnable" equilibrium? Why should we care about such a threat?

2. Model setup

For space and clarity, I will concentrate entirely on the simple model, presented on McCallum's pp. 4–5 with learnability results on p. 12, and I will use McCallum's notation. To recap, the model consists of a "Fisher equation" coming from consumer's first-order conditions:

$$R_t = E_t \pi_{t+1},$$

and a Taylor rule, expressing the central bank's behavior:

$$R_t = \frac{1}{a}\pi_t + e_t,$$

$$e_t = \rho e_{t-1} + \varepsilon_t. \tag{1}$$

(I leave out constants in both equations.) Eliminating the nominal interest rate R_t, the equilibrium condition is

$$\pi_t = aE_t\pi_{t+1} + u_t, \tag{2}$$

where

$$u_t \equiv -ae_t.$$

Assume that $1/a > 1$, following the Taylor principle. The equilibria of this model are any inflation path that solves the difference equation:

$$\pi_{t+1} = \frac{1}{a}\pi_t + e_t + \delta_{t+1}, \tag{3}$$

where δ_{t+1} is *any* random variable with $E_t\delta_{t+1} = 0$. This arbitrariness of δ indexes the multiplicity of equilibrium inflation paths.

It is useful to rewrite this equilibrium condition as

$$(\pi_{t+1} - \psi u_{t+1}) = \frac{1}{a}(\pi_t - \psi u_t) + a\psi\varepsilon_{t+1} + \delta_{t+1} \tag{4}$$

where

$$\psi \equiv \frac{1}{1-\rho a}.$$

(From (3), subtract ψu_{t+1} from both sides, add and subtract $(1/a)\psi u_t$ from the right hand side, and simplify.) Now it is easy to see that there is a "unique locally bounded" equilibrium

$$\pi_t = \psi u_t. \tag{5}$$

Equivalently, in this equilibrium the variable δ_t that indexes equilibria jumps around with the monetary policy shock as

$$\delta_t = -a\psi\varepsilon_t. \tag{6}$$

Given a choice of equilibrium at date t, expected future inflation follows:

$$E_t\pi_{t+j} - \psi\rho^{j+1}u_t = \frac{1}{a^j}(\pi_t - \psi u_t).$$

Thus, equilibrium (5)–(6) is the only one not expected to explode (unique locally bounded). But, as we agree, nothing yet rules out nominal explosions.

3. McCallum: the locally bounded equilibrium is learnable

Now, suppose agents do not know ψ and instead "estimate the relationship $\pi_t = \psi_{1t}u_t$." (ψ_1 means estimated, ψ_{1t} means the estimated value at time t.)

The simplest "learnability" criterion is this: If you woke up in the equilibrium $\pi_t = \psi u_t$, and you did not know ψ, could you (one individual) learn it from data on ψ_t and u_t? In this case, there is no error term in the regression. You would know ψ after one data point. I will call this circumstance an "individually learnable" equilibrium, and that is certainly true here.

McCallum and Evans and Honkapohja study "convergence," a more subtle concept of learnability. Here, we suppose that *nobody* knows the parameter ψ, but they guess it and update. When nobody knows ψ, the economy will potentially be in a different equilibrium, and everybody might "learn" the wrong lessons, in such a way never to end up at the right equilibrium. The ability of an equilibrium to be individually learnable is (I think) necessary for convergence, but not sufficient.

This example also converges, as follows. If people start with a guess ψ_{1t}, and the Fed just follows the usual rule (1), then from (2), equilibrium inflation is

$$\pi_t = aE_t\pi_{t+1} + u_t = a\psi_{1t}\rho u_t + u_t = (1+\rho a\psi_{1t})u_t.$$

In addition, suppose people update slowly, waiting a period to form a new guess from the perfectly measured relation they just observed,

$$\psi_{1t+1} = 1 + \rho a\psi_{1t}. \tag{7}$$

Iterating forward, this sequence of guesses leads to

$$\lim_{t\to\infty}\psi_{1t} = \psi.$$

Though there are smarter things people could do, which converge faster, this establishes that agents can learn with very simple rules.

(McCallum assumes that people run linear regressions. He establishes convergence with least-squares learning by applying Evans and Honkapohja's results, i.e. "Clearly the learnability conditions analogous to (15a)–(15c) are that the 1×1 matrices a, ρ, and $a\rho$ all have eigenvalues with real parts less than 1." The result is the same, but cannot be reproduced in the compact manner of my example.)

McCallum also claims that the non-locally bounded equilibria are *not* learnable in "analogous fashion." This is important, of course: if all equilibria are learnable, then learnability does not help us to narrow them down. Though I disagree with the result, there are more important issues to settle first.

4. Observable shocks and the opposite result

There is a deeper problem with McCallum's approach: *He assumes that u_t is directly observable*. To run a regression of π_t on u_t, you need data on u_t! $u_t = -e_t/a$ is the monetary policy shock. We do not usually think that agents can observe it directly.

Can't agents learn the monetary policy shock by running regressions, i.e. can't they infer e_t from observations of R_t and π_t and the policy rule

$$R_t = \frac{1}{a}\pi_t + e_t? \tag{8}$$

Alas, no. To measure e_t in this way, you need to know a. And agents cannot learn a.

If they try to run Eq. (8) as a regression, they fail because the right hand variable is perfectly correlated with the error term: The bounded-equilibrium value of π_t is

$$\pi_t = \psi u_t = -\frac{\psi}{a}e_t.$$

More generally, the observables in this model are (R_t, π_t). Since u_t follows an AR(1) and the bounded equilibrium is $\pi_t = \psi u_t$, the observables in the bounded equilibrium follow the laws of motion

$$\pi_t = \rho\pi_{t-1} + v_t, \tag{9}$$

$$R_t = \rho\pi_t, \tag{10}$$

where v_t is a regression error. *This* is what agents can see and learn. And *the parameter a does not appear. a is not identified from the equilibrium dynamics of the bounded equilibrium.* (This is the major point of Cochrane, 2007.) There is absolutely no way for agents to learn a even individually, and hence there is no way for them to measure u_t. Eq. (10) shows that the regression of R_t on π_t measures ρ, not $1/a$.

In sum, to run McCallum's "learnability" regression of π_t on u_t, agents need to be given, "from introspection or divine revelation," direct observation of the monetary policy shock e_t or prior knowledge of the parameter a. Without that assumption, I obtain the opposite result: *the bounded equilibrium is not learnable.*

As another way to see the point, suppose that the Fed follows a policy that violates the Taylor principle,

$$R_t = \rho\pi_t, \quad \rho < 1.$$

Now, the equilibrium dynamics are

$$\pi_{t+1} = \rho\pi_t + \delta_{t+1}.$$

There are multiple bounded equilibria this time. But the equilibrium dynamics (9) and (10) are *exactly* the same as in this case. There is no way for an agent to learn whether he lives in the Taylor rule, unique bounded equilibrium world with $1/a > 1$ or in this multiple equilibrium, indeterminate world, by running any regressions.

The converse result holds for the explosive solutions. These solutions follow equilibrium dynamics

$$\pi_{t+1} = \frac{1}{a}\pi_t + e_t + \delta_{t+1}.$$

As before, π_t and e_t are potentially correlated, so an OLS regression of π_{t+1} on π_t is biased in a finite sample. However, now π_t is an explosive process, while e_t and δ_{t+1} are stationary. Thus, the regression coefficient of π_{t+1} on π_t converges to $1/a$! *The explosive solutions are individually learnable.* I do not know if the explosive solutions converge, but neither I nor McCallum cares about this extra step—we do not want these anyway, though for different reasons.

In sum, once we remove the assumption that agents can see the monetary policy shock directly, I reverse McCallum's main claimed results: the bounded equilibrium is *not* learnable and the explosive equilibria *are* learnable (at least individually).

5. Equilibrium dynamics: a subtle trap

Now, agents can learn the law of motion (9) of the bounded equilibrium and hence make correct forecasts $E_t\pi_{t+j} = \rho^j\pi_t$. It is easy to conclude that this equilibrium is therefore learnable, despite the fact that agents cannot learn a or observe u_t.

This is a mistake. The whole point of this class of models is that knowledge of the bounded-equilibrium dynamics is not enough. Agents must know that alternative equilibria will lead to explosions. Equivalently, they must know that only one value of inflation today π_t is consistent with "anchored" long run expectations $\lim_{j\to\infty} E_t\pi_{t+j}$. They must know they are in the Taylor world with $1/a > 1$ not the observationally equivalent world with $1/a = \rho < 1$. If agents form expectations based on the equilibrium law of motion $E_t\pi_{t+j} = \rho^j\pi_t$ then *any* value of inflation today can be an equilibrium. Much of Woodford's (2003) *Interest and Prices* is dedicated to this point. So much new-Keynesian writing is devoted to communicating (by means other than regressions) the Fed's commitments to the Taylor rule that I do not think it is necessary to go on.

I think more generally this is where McCallum's analysis goes wrong. When discussing learnability in general, in section 3, he sets up a system with endogenous variables x and exogenous variables z, which follow

$$x_t = AE_t x_{t+1} + Cx_{t-1} + Dz_t,$$

$$z_t = Rz_{t-1} + \varepsilon_t.$$

(These are McCallum's Eqs. (10) and (11).) He then considers "fundamental solutions of form

$$x_t = \Omega x_{t-1} + \Gamma z_t$$

and he writes "for RE to prevail, agents need to have their expectations based on accurate quantitative knowledge of Ω and Γ, what the agents need to learn about is the system's [equilibrium] law of motion."

That is *not* enough in a new-Keynesian model. For the new-Keynesian equilibrium to form and be locally unique, agents must know A as well, and they cannot learn that quantity from the bounded equilibrium of this model.

McCallum could really have helped us by showing how learning works, and how equilibria are formed, in the simple new-Keynesian model. When agents forecast based on past values of endogenous variables, all the forward-looking components so vital to the bounded equilibrium of the new-Keynesian model vanish. Yet such forecasts are central to McCallum's analysis.

For example, let us follow McCallum's analysis of convergence in the non-fundamentals solution (where he writes " by contrast, the non-fundamentals solution (9) yields...). McCallum studies explosive equilibria of the form (his (8) and (9)),

$$\pi_t = \frac{1}{\rho}e_t + \frac{1}{a}\pi_{t-1}. \tag{11}$$

As before, we suppose people have a guess of this form,

$$\pi_t = \psi_{1t}u_t + \psi_{2t}\pi_{t-1}.$$

What equilibrium results? If we plug such forecasts into the equilibrium condition (2), we obtain

$$\pi_t = (1 + \rho a\psi_{1t})u_t + a\psi_{2t}\pi_t.$$

Since people forecast π_{t+1} based on the endogenous variable π_t, it appears on both sides. We could look for an "equilibrium" here,

$$\psi_t = \frac{1}{1 - a\psi_{2t}}(1 + \rho a\psi_{1t})u_t.$$

This expression is not even of the same form as the equilibrium (11). (The "fundamentals" solution avoided this problem by ruling π_{t-1} out of agents' regression a-priori. Of course, somehow agents needed to know they were not in the non-fundamentals equilibrium to know they should do that.) Perhaps this is what McCallum means by the statement that the non-fundamentals solution is not learnable. But it seems clearly to fall into the above trap—agents cannot forecast based only on endogenous variables in the new-Keyensian model. And it deeply begs the question, how does the "equilibrium" of a new-Keynesian model work, in which agents are forecasting based on the same π_t that is being determined? How is an equilibrium sensibly formed when agents instantaneously update expectations as the endogenous variable is determined?

Clarifying all this would be an enormous help. Alas, McCallum just maps into Evans and Honkapohja's general framework, and finds the eigenvalues of matrices. But it is not clear how or even if that framework applies to this problem.

6. Even so, would it work?

Even if McCallum's learnability results (and those of related authors, such as Woodford, 2003) were correct, I do not think they would solve the fundamental problem of multiple equilibria in new-Keynesian models.

How does the Taylor rule work? The Fed, undergraduate textbooks, and everyone's op-eds say that it *stabilizes* the economy, introducing an eigenvalue *less than* one. If inflation rises, the Fed raises interest rates, and this lowers "demand," which lowers future inflation.

The standard new-Keynesian model logic, as presented here, works in the opposite way. The Taylor rule *destabilizes* the economy. If inflation rises, the Fed commits to raise future inflation, and leads us off to nominal explosions. Ruling out such explosions, inflation must be where the Fed wants it to be in the first place. As McCallum and I agree, the "ruling this out" part stands on shaky ground. Nothing in economics rules out such a hyperinflation, and words such as Woodford (2003, p. 128) are hardly reassuring:

> The equilibrium ... is nonetheless *locally* unique, which may be enough to allow expectations to coordinate upon that equilibrium rather than on one of the others.

I think the non-learnability of the *bounded* solution puts it on shakier ground still—there is no way for agents to learn of the Fed's hyperinflationary threats from time-series observations. They would have to learn it from Fed pronouncements—and the Fed is pronouncing exactly the opposite response. (Conversely, the learnability of the *unbounded* solutions seems pretty obvious—if you observe a Fed-induced hyperinflation, you can learn the Fed's reaction speed pretty quickly.)

But suppose I am wrong—suppose the bounded solution *is* learnable and the unbounded ones are not. McCallum is then saying that inflation is determinate at one particular value because the Fed, by explosive reaction, threatens for any other initial value to take us to an equilibrium that is "unlearnable." This does not sound like a threat that will strike fear in Wall Street's heart, and "coordinate expectations" on anything other than confusion. In the centuries-long search for the price-level anchor in a fiat-money economy, is the threat to take us to an "unlearnable" equilibrium for all but one initial value really the answer?

Acknowledgments

I thank Larry Christiano, George Evans, and Bennett McCallum for comments.

References

Cochrane, J.H., 2007. Identification with Taylor rules: a critical review. Manuscript, University of Chicago.
Evans, G.W., Honkapohja, S., 2001. Learning and Expectations in Macroeconomics. Princeton University Press, Princeton.
McCallum, B.T., 2009. Inflation determination with Taylor rules: is the new-Keynesian analysis critically flawed? Journal of Monetary Economics, in press, doi:10.1016/j.jmoneco.2009.10.003.
Woodford, M., 2003. Interest and Prices. Princeton University Press, Princeton.