

The Trouble With Macroeconomics

PAUL ROMER

Stern School of Business
New York University

Wednesday 14th September, 2016

Abstract

For more than three decades, macroeconomics has gone backwards. The treatment of identification now is no more credible than in the early 1970s but escapes challenge because it is so much more opaque. Macroeconomic theorists dismiss mere facts by feigning an obtuse ignorance about such simple assertions as "tight monetary policy can cause a recession." Their models attribute fluctuations in aggregate variables to imaginary causal forces that are not influenced by the action that any person takes. A parallel with string theory from physics hints at a general failure mode of science that is triggered when respect for highly regarded leaders evolves into a deference to authority that displaces objective fact from its position as the ultimate determinant of scientific truth.

Delivered January 5, 2016 as the Commons Memorial Lecture of the Omicron Delta Epsilon Society. Forthcoming in *The American Economist*.

Lee Smolin begins *The Trouble with Physics* (Smolin 2007) by noting that his career spanned the only quarter-century in the history of physics when the field made no progress on its core problems. The trouble with macroeconomics is worse. I have observed more than three decades of intellectual regress.

In the 1960s and early 1970s, many macroeconomists were cavalier about the identification problem. They did not recognize how difficult it is to make reliable inferences about causality from observations on variables that are part of a simultaneous system. By the late 1970s, macroeconomists understood how serious this issue is, but as Canova and Sala (2009) signal with the title of a recent paper, we are now "Back to Square One." Macro models now use incredible identifying assumptions to reach bewildering conclusions. To appreciate how strange these conclusions can be, consider this observation, from a paper published in 2010, by a leading macroeconomist:

... although in the interest of disclosure, I must admit that I am myself less than totally convinced of the importance of money outside the case of large inflations.

1 Facts

If you want a clean test of the claim that monetary policy does not matter, the Volcker deflation is the episode to consider. Recall that the Federal Reserve has direct control over the monetary base, which is equal to currency plus bank reserves. The Fed can change the base by buying or selling securities.

Figure 1 plots annual data on the monetary base and the consumer price index (CPI) for roughly 20 years on either side of the Volcker deflation. The solid line in the top panel (blue if you read this online) plots the base. The dashed (red) line just below is the CPI. They are both defined to start at 1 in 1960 and plotted on a ratio scale so that moving up one grid line means an increase by a factor of 2. Because of the ratio scale, the rate of inflation is the slope of the CPI curve.

The bottom panel allows a more detailed year-by-year look at the inflation rate, which is plotted using long dashes. The straight dashed lines show the fit of a linear trend to the rate of inflation before and after the Volcker deflation. Both panels use shaded regions to show the NBER dates for business cycle contractions. I highlighted the two recessions of the Volcker deflation with darker shading. In both the upper and lower panels, it is obvious that the level and trend of inflation changed abruptly around the time of these two recessions.

When one bank borrows reserves from another, it pays the nominal federal funds rate. If the Fed makes reserves scarce, this rate goes up. The best indicator of

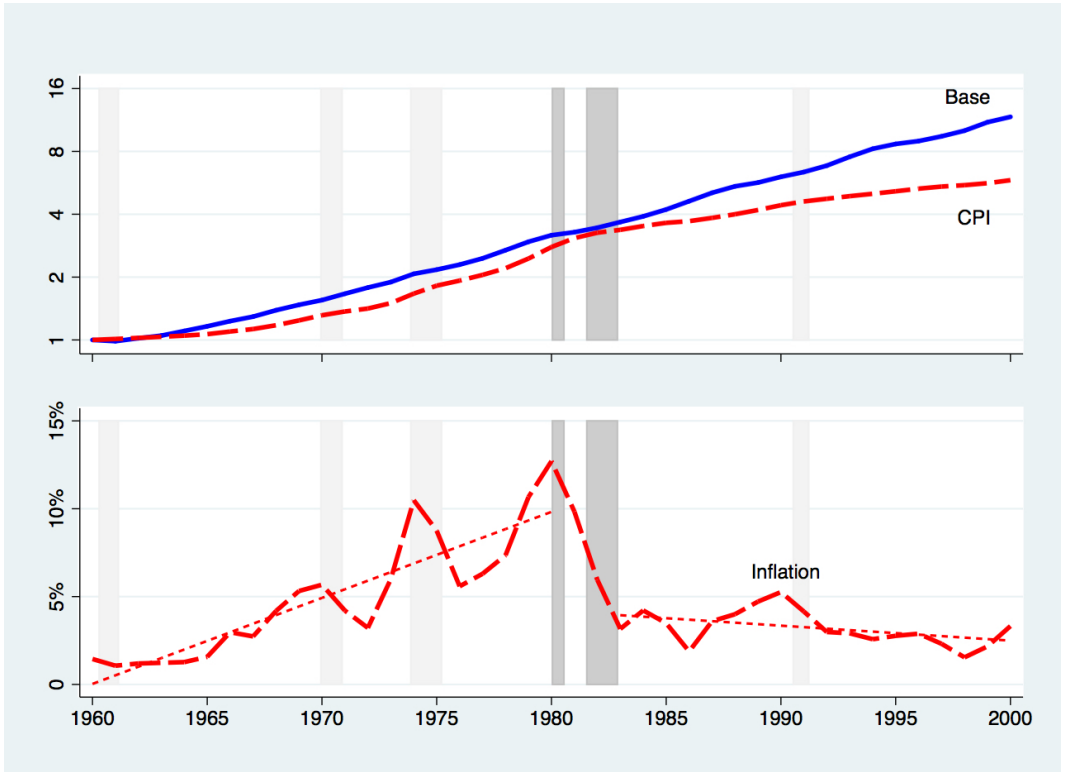


Figure 1: Money and Prices from 1960 to 2000

monetary policy is the real federal funds rate – the nominal rate minus the inflation rate. This real rate was higher during Volcker’s term as Chairman of the Fed than at any other time in the post-war era.

Two months into his term, Volcker took the unusual step of holding a press conference to announce changes that the Fed would adopt in its operating procedures. Romer and Romer (1989) summarize the internal discussion at the Fed that led up to this change. Fed officials expected that the change would cause a "prompt increase in the Fed Funds rate" and would "dampen inflationary forces in the economy."

Figure 2 measures time relative to August 1979, when Volcker took office. The solid line (blue if you are seeing this online) shows the increase in the real fed funds rate, from roughly zero to about 5%, that followed soon thereafter. It subtracts from the nominal rate the measure of inflation displayed as the dotted (red) line. It plots inflation each month, calculated as the percentage change in the CPI over the previous 12 months. The dashed (black) line shows the unemployment rate, which in contrast to GDP, is available on a monthly basis. During the first recession, output fell by 2.2% as unemployment rose from 6.3% to 7.8%. During the second, output fell by

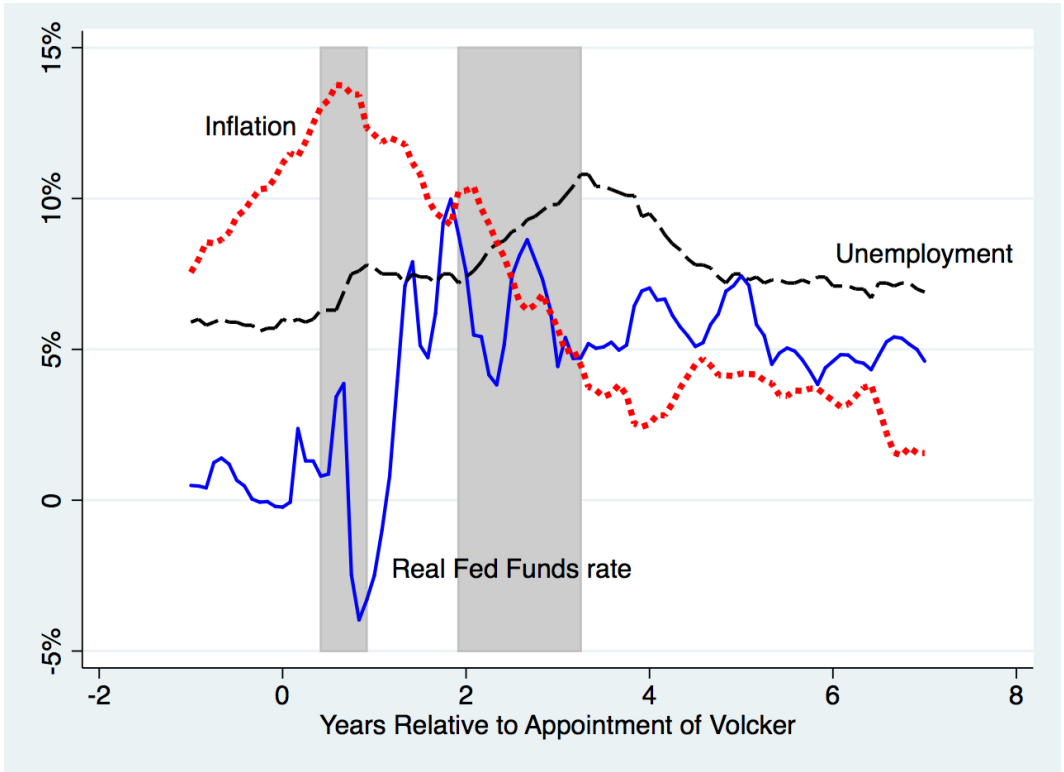


Figure 2: The Volker Deflation

2.9% as unemployment increased from 7.2% to 10.8%.

The data displayed in Figure 2 suggest a simple causal explanation for the events that is consistent with what the Fed insiders predicted:

1. The Fed aimed for a nominal Fed Funds rate that was roughly 500 basis points higher than the prevailing inflation rate, departing from this goal only during the first recession.
2. High real interest rates decreased output and increased unemployment.
3. The rate of inflation fell, either because the combination of higher unemployment and a bigger output gap caused it to fall or because the Fed's actions changed expectations.

If the Fed can cause a 500 basis point change in interest rates, it is absurd to wonder if monetary policy is important. Faced with the data in Figure 2, the only way to remain faithful to dogma that monetary policy is not important is to argue that despite what people at the Fed thought, they did not change the Fed funds rate; it

was an imaginary shock that increased it at just the right time and by just the right amount to fool people at the Fed into thinking they were the ones who were the ones moving it around.

To my knowledge, no economist will state as fact that it was an imaginary shock that raised real rates during Volcker's term, but many endorse models that will say this for them.

2 Post-Real Models

Macroeconomists got comfortable with the idea that fluctuations in macroeconomic aggregates are caused by imaginary shocks, instead of actions that people take, after Kydland and Prescott (1982) launched the real business cycle (RBC) model. Stripped to its essentials, an RBC model relies on two identities. The first defines the usual growth-accounting residual as the difference between the growth of output Y and growth of an index X of inputs in production:

$$\Delta\%A = \Delta\%Y - \Delta\%X$$

Abromovitz (1956) famously referred to this residual as "the measure of our ignorance." In his honor and to remind ourselves of our ignorance, I will refer to the variable A as phlogiston.

The second identity, the quantity theory of money, defines velocity v as nominal output (real output Y times the price level P) divided by the quantity of a monetary aggregate M :

$$v = \frac{YP}{M}$$

The real business cycle model explains recessions as exogenous decreases in phlogiston. Given output Y , the only effect of a change in the monetary aggregate M is a proportional change in the price level P . In this model, the effects of monetary policy are so insignificant that, as Prescott taught graduate students at the University of Minnesota "postal economics is more central to understanding the economy than monetary economics" (Chong, La Porta, Lopez-de-Silanes, Shliefer, 2014).

Proponents of the RBC model cite its microeconomic foundation as one of its main advantages. This posed a problem because there is no microeconomic evidence for the negative phlogiston shocks that the model invokes nor any sensible theoretical interpretation of what a negative phlogiston shock would mean.

In private correspondence, someone who overlapped with Prescott at the University of Minnesota sent me an account that helped me remember what it was like to encounter "negative technology shocks" before we all grew numb:

I was invited by Ed Prescott to serve as a second examiner at one of his student's preliminary oral. ... I had not seen or suspected the existence of anything like the sort of calibration exercise the student was working on. There were lots of reasons I thought it didn't have much, if any, scientific value, but as the presentation went on I made some effort to sort them out, and when it came time for me to ask a question, I said (not even imagining that the concept wasn't specific to this particular piece of thesis research) "What are these technology shocks?"

Ed tensed up physically like he had just taken a bullet. After a very awkward four or five seconds he growled "They're that traffic out there." (We were in a room with a view of some afternoon congestion on a bridge that collapsed a couple decades later.) Obviously, if he had what he sincerely thought was a valid justification of the concept, I would have been listening to it ...

What is particularly revealing about this quote is that it shows that if anyone had taken a micro foundation seriously it would have put a halt to all this lazy theorizing. Suppose an economist thought that traffic congestion is a metaphor for macro fluctuations or a literal cause of such fluctuations. The obvious way to proceed would be to recognize that drivers make decisions about when to drive and how to drive. From the interaction of these decisions, seemingly random aggregate fluctuations in traffic throughput will emerge. This is a sensible way to think about a fluctuation. It is totally antithetical to an approach that assumes the existence of imaginary traffic shocks that no person does anything to cause.

In response to the observation that the shocks are imaginary, a standard defense invokes Milton Friedman's (1953) methodological assertion from unnamed authority that "the more significant the theory, the more unrealistic the assumptions (p.14)." More recently, "all models are false" seems to have become the universal hand-wave for dismissing any fact that does not conform to the model that is the current favorite.

The noncommittal relationship with the truth revealed by these methodological evasions and the "less than totally convinced ..." dismissal of fact goes so far beyond post-modern irony that it deserves its own label. I suggest "post-real."

3 DSGE Extensions to the RBC Core

3.1 More imaginary shocks

Once macroeconomists concluded that it was reasonable to invoke an imaginary forcing variables, they added more. The resulting menagerie, together with my

suggested names now includes:

- A general type of phlogiston that increases the quantity of consumption goods produced by given inputs
- An "investment-specific" type of phlogiston that increases the quantity of capital goods produced by given inputs
- A troll who makes random changes to the wages paid to all workers
- A gremlin who makes random changes to the price of output
- Aether, which increases the risk preference of investors
- Caloric, which makes people want less leisure

With the possible exception of phlogiston, the modelers assumed that there is no way to directly measure these forces. Phlogiston can in measured by growth accounting, at least in principle. In practice, the calculated residual is very sensitive to mismeasurement of the utilization rate of inputs, so even in this case, direct measurements are frequently ignored.

3.2 Sticky Prices

To allow for the possibility that monetary policy could matter, empirical DSGE models put sticky-price lipstick on this RBC pig. The sticky-price extensions allow for the possibility that monetary policy can affect output, but the reported results from fitted or calibrated models never stray far from RBC dogma. If monetary policy matters at all, it matters very little.

As I will show later, when the number of variables in a model increases, the identification problem gets much worse. In practice, this means that the econometrician has more flexibility in determining the results that emerge when she estimates the model.

The identification problem means that to get results, an econometrician has to feed in something other than data on the variables in the simultaneous system. I will refer to things that get fed in as facts with unknown truth value (FWUTV) to emphasize that although the estimation process treats the FWUTV's as if they were facts known to be true, the process of estimating the model reveals nothing about the actual truth value. The current practice in DSGE econometrics is feed in some FWUTV's by "calibrating" the values of some parameters and to feed in others tight Bayesian priors. As Olivier Blanchard (2016) observes with his typical understatement, "in many cases, the justification for the tight prior is weak at best,

and what is estimated reflects more the prior of the researcher than the likelihood function."

This is more problematic than it sounds. The prior specified for one parameter can have a decisive influence on the results for others. This means that the econometrician can search for priors on seemingly unimportant parameters to find ones that yield the expected result for the parameters of interest.

3.3 An Example

The Smets and Wouters (SW) model was hailed as a breakthrough success for DSGE econometrics. When they applied this model to data from the United States for years that include the Volcker deflation, Smets and Wouters (2007) conclude:

...monetary policy shocks contribute only a small fraction of the forecast variance of output at all horizons (p. 599).

...monetary policy shocks account for only a small fraction of the inflation volatility (p. 599).

...[In explaining the correlation between output and inflation:] Monetary policy shocks do not play a role for two reasons. First, they account for only a small fraction of inflation and output developments (p. 601).

What matters in the model is not money but the imaginary forces. Here is what the authors say about them, modified only with insertions in bold and the abbreviation "AKA" as a stand in for "also known as."

While "demand" shocks such as the **aether AKA** risk premium, exogenous spending, and investment-specific **phlogiston AKA** technology shocks explain a significant fraction of the short-run forecast variance in output, both **the troll's** wage mark-up (or **caloric AKA** labor supply) and, to a lesser extent, output-specific **phlogiston AKA** technology shocks explain most of its variation in the medium to long run. ... Third, inflation developments are mostly driven by the **gremlin's** price mark-up shocks in the short run and the **troll's** wage mark-up shocks in the long run (p. 587).

A comment in a subsequent paper (Linde, Smets, Wouters 2016, footnote 16) underlines the flexibility that imaginary driving forces bring to post-real macroeconomics (once again with my additions in bold):

The prominent role of the **gremlin's** price and **the troll's** wage markup for explaining inflation and behavior of real wages in the SW-model have

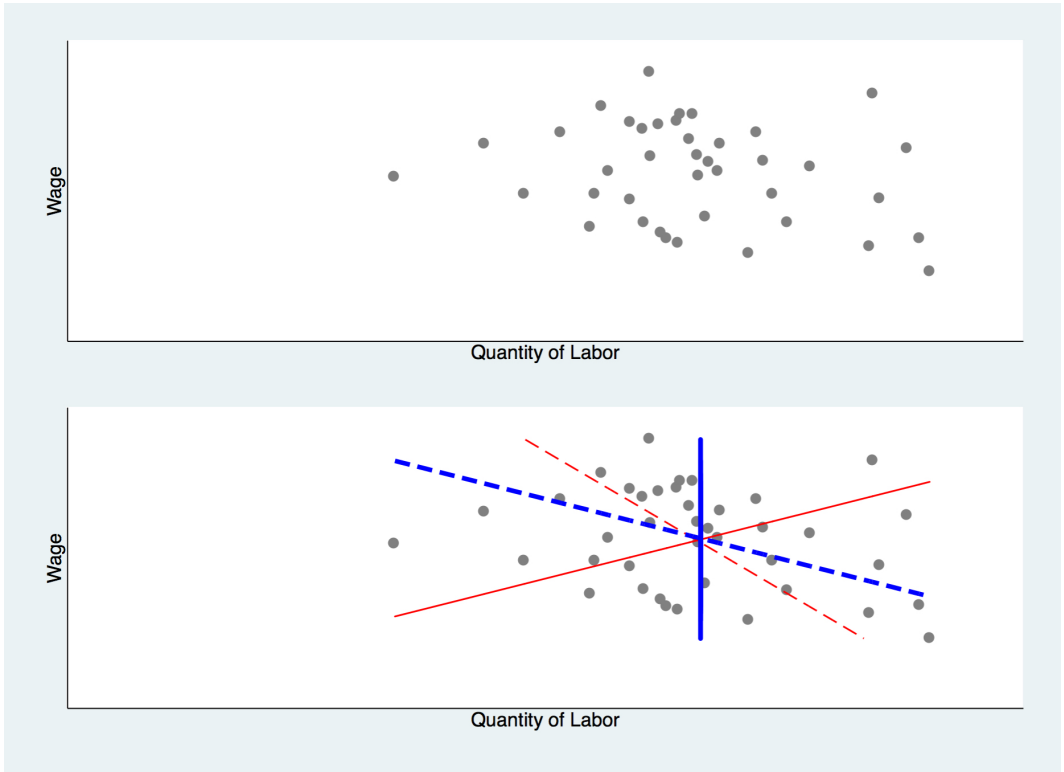


Figure 3: A Hypothetical Labor Market

been criticized by Chari, Kehoe and McGrattan (2009) as implausibly large. Galí, Smets and Wouters (2011), however, shows that the size of the markup shocks can be reduced substantially by allowing for **caloric** AKA preference shocks to household preferences.

4 The Identification Problem

A modeling strategy that allows imaginary shocks and hence more variables makes the identification problem worse. This offers more flexibility in determining how the results from of any empirical exercise turn out.

4.1 Identification of the Simplest Supply and Demand Model

The way to think about any question involving identification is to start by posing it in a market with a supply curve and a demand curve. Suppose we have data like those in Figure 3 on (the log of) wages w and (the log of) hours worked ℓ . To predict the

effect of a policy change, economists need to know the elasticity of labor demand. Here, the identification problem means that there is no way to calculate this elasticity from the scatter plot alone.

To produce the data points in the figure, I specified a data generating process with a demand curve and a supply curve that are linear in logs plus some random shocks. Then I tried to estimate the underlying curves using only the data. I specified a model with linear supply and demand curves and independent errors and asked my statistical package to calculate the two intercepts and two slopes. The software package barfed. (Software engineers assure me, with a straight face, this is the technical term for throwing an error.)

Next, I fed in a fact with an unknown truth value (a FWUTV) by imposing the restriction that the supply curve is vertical. (To be more explicit, the truth value of this restriction is unknown to you because I have not told you what I know to be true about the curves that I used to generate the data.) With this FWUTV, the software returned the estimates illustrated by the thick lines (blue if you are viewing this paper online) in the lower panel of the figure. The accepted usage seems to be that one says "the model is identified" if the software does not barf.

Next, I fed in a different FWUTV by imposing the restriction that the supply curve passes through the origin. Once again, the model is identified; the software did not barf. It produced as output the parameters for the thin (red) lines in the lower panel.

You do not know if either of these FWUTV's is true, but you know that at least one of them has to be false and nothing about the estimates tells you which it might be. So in the absence of any additional information, the elasticity of demand produced by each of these identified-in-the-sense-that-the-software-does-not-barf models is meaningless.

4.2 The m^2 Scaling Problem

Suppose that x is a vector of observations on m variables. Write a linear simultaneous equation model of their interaction as

$$x = Sx + c + \epsilon.$$

where the matrix S has zeros on its diagonal so that equation says that each component of x is equal to a linear combination of the other components plus a constant and an error term. For simplicity, assume that based on some other source of information, we know that the error ϵ_t is an independent draw in each period. Assume as well that none of the components of x is a lagged value of some other component. This equation has m^2 parameters to estimate because the matrix S has $m(m - 1)$ off-diagonal slope parameters and there are m elements in the constant c .

The error terms in this system could include omitted variables that influence several of the observed variables, so there is no a priori basis for assuming that the errors for different variables in the list x are uncorrelated. (The assumption of uncorrelated error terms for the supply curve and the demand curve was another FWUTV that I snuck into the estimation processes that generated the curves in the bottom panel of Figure 3.) This means that all of the information in the sample estimate of the variance-covariance matrix of the x 's has to be used to calculate the nuisance parameters that characterize the variance-covariance matrix of the ϵ 's.

So this system has m^2 parameters to calculate from only m equations, the ones that equate $\mu(x)$, the expected value of x from the model, to \bar{x} , the average value of x observed in the data.

$$\bar{x} = S\mu(x) + c.$$

The Smets-Wouters model, which has 7 variables, has $7^2 = 49$ parameters to estimate and only 7 equations, so 42 FWUTV's have to be fed in to keep the software from barfing.

4.3 Adding Expectations Makes the Identification Problem Twice as Bad

In their critique of traditional Keynesian models, Lucas and Sargent (1979) seem to suggest that rational expectations will help solve the identification problem by introducing a new set of "cross-equation restrictions."

To see what happens when expectations influence decisions, suppose that the expected wage has an effect on labor supply that is independent from the spot wage because people use the expected wage to decide whether to go to the spot market. To capture this effect, the labor supply equation must include a term that depends on the expected wage $\mu(w)$.

Generalizing, we can add to the previous linear system another $m \times m$ matrix of parameters B that captures the effect of $\mu(x)$:

$$x = Sx + B\mu(x) + c + \epsilon. \quad (1)$$

This leaves a slightly different set of m equations to match up with the average value \bar{x} from the data:

$$\bar{x} = S\mu(x) + B\mu(x) + c \quad (2)$$

From these m equations, the challenge now is to calculate twice as many parameters, $2m^2$. In a system with seven variables, this means $2 * 7^2 - 7 = 91$ parameters that have to be specified based on information other than what is in time series for x .

Moreover, in the absence of some change to a parameter or to the distribution of the errors, the expected value of x will be constant, so even with arbitrarily many observations on x and all the knowledge one could possibly hope for about the slope coefficients in S , it will still be impossible to disentangle the constant term c from the expression $B\mu(x)$.

So allowing for the possibility that expectations influence behavior makes the identification problem at least twice as bad. This may be part of what Sims (1980) had in mind when he wrote, "It is my view, however, that rational expectations is more deeply subversive of identification than has yet been recognized." Sim's paper, which is every bit as relevant today as it was in 1980, also notes the problem from the previous section, that the number of parameters that need to be pinned down scale as the square of the number of variables in the model; and it attributes the impossibility of separating the constant term from the expectations term to Solow (1974).

5 Regress in the Treatment of Identification

Post-real macroeconomists have not delivered the careful attention to the identification problem that Lucas and Sargent (1979) promised. They still rely on FWUTV's. All they seem to have done is find new ways to feed in FWUTV's.

5.1 Natural Experiments

Faced with the challenge of estimating the elasticity of labor demand in a supply and demand market, the method of Friedman and Schwartz (1963) would be to look for two periods that are adjacent in time, with conditions that were very similar, except for a change that shifts the labor supply curve in one period relative to the other. To find this pair, they would look carefully at historical evidence that they would add to the information in the scatter plot.

If the historical circumstances offer up just one such pair, they would ignore all the other data points and base an estimate on just that pair. If Lucas and Sargent (1979) are correct that the identification problem is the most important problem in empirical macroeconomics, it makes sense to throw away data. It is better to have a meaningful estimate with a larger standard error than a meaningless estimate with a small standard error.

The Friedman and Schwartz approach feeds in a fact with a truth value that others can assess. This allows cumulative scientific analysis of the evidence. Of course, allowing cumulative scientific analysis means opening your results up to criticism and revision.

When I was in graduate school, I was impressed by the Friedman Schwartz account of the increase in reserve requirements that, they claimed, caused the sharp recession of 1938-9. Romer and Romer (1989) challenge this reading of the history and of several other episodes from the Great Depression. They suggest that the most reliable identifying information comes from the post-war period, especially the Volcker deflation. Now my estimate of the effect that monetary policy can have on output in the United States relies much more heavily on the cleanest experiment – the Volcker deflation.

5.2 Identification by Assumption

As Keynesian macroeconomic modelers increased the number of variables that they included, they ran smack into the problem of m^2 parameters to estimate from m equations. They responded by feeding in as FWUTV's the values for many of the parameters, mainly by setting them equal to zero. As Lucas and Sargent note, in many cases there was no independent evidence one could examine to assess the truth value of these FWUTV's. But to their credit, the Keynesian model builders were transparent about what they did.

5.3 Identification by Deduction

A key part of the solution to the identification problem that Lucas and Sargent (1979) seemed to offer was that mathematical deduction could pin down some parameters in a simultaneous system. But solving the identification problem means feeding facts with truth values that can be assessed, yet math cannot establish the truth value of a fact. Never has. Never will.

In practice, what math does is let macroeconomists locate the FWUTV's farther away from the discussion of identification. The Keynesians tended to say "Assume P is true. Then the model is identified." Relying on a micro-foundation lets an author can say, "Assume A, assume B, ... blah blah blah And so we have proven that P is true. Then the model is identified."

To illustrate this process in the context of the labor market example with just enough "blah blah" to show how this goes, imagine that a representative agent gets utility from consuming output $U(Y) = \frac{Y^\beta}{\beta}$ with $\beta < 1$ and disutility from work $-\gamma V(L) = -\gamma \frac{L^\alpha}{\alpha}$ with $\alpha > 1$. The disutility from labor depends on fluctuations in the level of aether captured by the random variable γ .

The production technology for output $Y = \pi AL$ is the product of labor times the

prevailing level of phlogiston, π , and a constant A . The social planner's problem is

$$\begin{aligned} \max \quad & U(Y) - \gamma * V(L) \\ \text{s.t.} \quad & Y = \pi * A * L \end{aligned}$$

To derive a labor supply curve and a labor demand curve, split this into two separate maximization problems that are connected by the wage W :

$$\max_{L_S, L_D} \quad \frac{(\pi A L_D)^\beta}{\beta} - W_D * L_D + W_S * L_S - \gamma \frac{L_S^\alpha}{\alpha}$$

Next, make some distributional assumptions about the imaginary random variables, γ and π . Specifically, assume that they are log normal, with $\log(\gamma) \sim N(0, \sigma_\gamma)$ and $\log(\pi) \sim N(0, \sigma_\pi)$. After a bit of algebra, the two first-order conditions for this maximization problem reduce to these simultaneous equations,

$$\begin{aligned} \text{Demand} \quad & \ell_D = a - b * w + \epsilon_D \\ \text{Supply} \quad & \ell_S = d * w + \epsilon_S \end{aligned}$$

where ℓ_D is the log of L_D , ℓ_S is the log of L_S and w is the log of the wage. This system has a standard, constant elasticity labor demand curve and, as if by an invisible hand, a labor supply curve with an intercept that is equal to zero.

With enough math, an author can be confident that most readers will never figure out where a FWUTV is buried. A discussant or referee cannot say that an identification assumption is not credible if they cannot figure out what it is and are too embarrassed to ask.

In this example, the FWUTV is that the mean of $\log(\gamma)$ is zero. Distributional assumptions about error terms are a good place to bury things because hardly anyone pays attention to them. Moreover, if a critic does see that this is the identifying assumption, how can she win an argument about the true expected value the level of aether? If the author can make up an imaginary variable, "because I say so" seems like a pretty convincing answer to any question about its properties.

5.4 Identification by Obfuscation

I never understood how identification was achieved in the current crop of empirical DSGE models. In part, they rely on the type of identification by deduction illustrated in the previous section. They also rely on calibration, which is the renamed version of identification by assumption. But I never knew if there were other places where FWUTV's were buried. The papers that report the results of these empirical exercises

do not discuss the identification problem. For example, in Smets and Wooters (2007), neither the word "identify" nor "identification" appear.

To replicate the results from that model, I read the User's Guide for the software package, Dynare, that the authors used. In listing the advantages of the Bayesian approach, the User's Guide says:

Third, the inclusion of priors also helps identifying parameters. (p. 78)

This was a revelation. Being a Bayesian means that your software never barfs.

In retrospect, this point should have been easy to see. To generate the thin curves in Figure 3, I used as a FWUTV the restriction that the intercept of the supply curve is zero. This is like putting a very tight prior on the intercept that is centered at zero. If I loosen up the prior a little bit and calculate a Bayesian estimate instead of a maximum likelihood estimate, I should get a value for the elasticity of demand that is almost the same.

If I do this, the Bayesian procedure will show that the posterior of the intercept for the supply curve is close to the prior distribution that I feed in. So in the jargon, I could say that "the data are not informative about the value of the intercept of the supply curve." But then I could say that "the slope of the demand curve has a tight posterior that is different from its prior." By omission, the reader could infer that it is the data, as captured in the likelihood function, that are informative about the elasticity of the demand curve when in fact it is the prior on the intercept of the supply curve that pins it down and yields a tight posterior. By changing the priors I feed in for the supply curve, I can change the posteriors I get out for the elasticity of demand until I get one I like.

It was news to me that priors are vectors for FWUTV's, but once I understood this and started reading carefully, I realized that was an open secret among econometricians. In the paper with the title that I note in the introduction, Canova and Sala (2009) write that "uncritical use of Bayesian methods, including employing prior distributions which do not truly reflect spread uncertainty, may hide identification pathologies." Onatski and Williams (2010) show that if you feed different priors into an earlier version of the Smets and Wooters model (2003), you get back different structural estimates. Iskrev (2010) and Komunjer and Ng (2011) note that without any information from the priors, the Smets and Wooter model is not identified. Reicher (2015) echos the point that Sims made in his discussion of the results of Hatanaka (1975). Baumeister and Hamilton (2015) note that in a bivariate vector autoregression for a supply and demand market that is estimated using Bayesian methods, it is quite possible that "even if one has available an infinite sample of data, any inference about the demand elasticity is coming exclusively from the prior distribution."

6 Questions About Economists, and Physicists

It helps to separate the standard questions of macroeconomics, such as whether the Fed can increase the real fed funds rate, from meta-questions about what economists do when they try to answer the standard questions. One example of a meta-question is why macroeconomists started invoking imaginary driving forces to explain fluctuations. Another is why they seemed to forget things that had been discovered about the identification problem.

I found that a more revealing meta-question to ask is why there are such striking parallels between the characteristics of string-theorists in particle physics and post-real macroeconomists. To illustrate the similarity, I will reproduce a list that Smolin (2007) presents in Chapter 16 of seven distinctive characteristics of string theorists:

1. Tremendous self-confidence
2. An unusually monolithic community
3. A sense of identification with the group akin to identification with a religious faith or political platform
4. A strong sense of the boundary between the group and other experts
5. A disregard for and disinterest in ideas, opinions, and work of experts who are not part of the group
6. A tendency to interpret evidence optimistically, to believe exaggerated or incomplete statements of results, and to disregard the possibility that the theory might be wrong
7. A lack of appreciation for the extent to which a research program ought to involve risk

The conjecture suggested by the parallel is that developments in both string theory and post-real macroeconomics illustrate a general failure mode of a scientific field that relies on mathematical theory. The conditions for failure are present when a few talented researchers come to be respected for genuine contributions on the cutting edge of mathematical modeling. Admiration evolves into deference to these leaders. Deference leads to effort along the specific lines that the leaders recommend. Because guidance from authority can align the efforts of many researchers, conformity to the facts is no longer needed as a coordinating device. As a result, if facts disconfirm the officially sanctioned theoretical vision, they are subordinated. Eventually, evidence

stops being relevant. Progress in the field is judged by the purity of its mathematical theories, as determined by the authorities.

One of the surprises in Smolin's account is his rejection of the excuse offered by the string theorists, that they do not pay attention to data because there is no practical way to collect data on energies at the scale that string theory considers. He makes a convincing case that there were plenty of unexplained facts that the theorists could have addressed if they had wanted to (Chapter 13). In physics as in macroeconomics, the disregard for facts has to be understood as a choice.

Smolin's argument lines up almost perfectly with a taxonomy for collective human effort proposed by Mario Bunge (1984). It starts by distinguishing "research" fields from "belief" fields. In research fields such as math, science, and technology, the pursuit of truth is the coordinating device. In belief fields such as religion and political action, authorities coordinate the efforts of group members.

There is nothing inherently bad about coordination by authorities. Sometimes there is no alternative. The abolitionist movement was a belief field that relied on authorities to make such decisions as whether its members should treat the incarceration of criminals as slavery. Some authority had to make this decision because there is no logical argument, nor any fact, that group members could use independently to resolve this question.

In Bunge's taxonomy, pseudoscience is a special type of belief field that claims to be science. It is dangerous because research fields are sustained by norms that are different from those of a belief field. Because norms spread through social interaction, pseudoscientists who mingle with scientists can undermine the norms that are required for science to survive. Revered individuals are unusually important in shaping the norms of a field, particularly in the role of teachers who bring new members into the field. For this reason, an efficient defense of science will hold the most highly regarded individuals to the highest standard of scientific conduct.

7 Loyalty Can Corrode The Norms of Science

This description of the failure mode of science should not be taken to mean that the threat to science arises when someone is motivated by self-interest. People are always motivated by self-interest. Science would never have survived if it required its participants to be selfless saints.

Like the market, science is a social system that uses competition to direct the self-interest of the individual to the advantage of the group. The problem is that competition in science, like competition in the market, is vulnerable to collusion.

Bob Lucas, Ed Prescott, and Tom Sargent led the development of post-real macroeconomics. Prior to 1980, they made important scientific contributions

to macroeconomic theory. They shared experience "in the foxhole" when these contributions provoked return fire that could be sarcastic, dismissive, and wrong-headed. As a result, they developed a bond of loyalty that would be admirable and productive in many social contexts.

Two examples illustrate the bias that loyalty can introduce into science.

7.1 Example 1: Lucas Supports Prescott

In his 2003 Presidential Address to the American Economics Association, Lucas gave a strong endorsement to Prescott's claim that monetary economics was a triviality.

This position is hard to reconcile with Lucas's 1995 Nobel lecture, which gives a nuanced discussion of the reasons for thinking that monetary policy does matter and the theoretical challenge that this poses for macroeconomic theory. It is also inconsistent with his comments (Lucas, 1994, p. 153) on a paper by Ball and Mankiw (1994), in which Lucas wrote that that Cochrane (1994) gives "an accurate view of how little can said to be firmly established about the importance and nature of the real effects of monetary instability, at least for the U.S. in the postwar period."

Cochrane notes that if money has the type of systematic effects that his VAR's suggest, it was more important to study such elements of monetary policy as the role of lender of last resort and such monetary institutions as deposit insurance than to make "an assessment of how much output can be further stabilized by making monetary policy more predictable." According to Cochrane (1994, p. 331), if this assessment suggests tiny benefits, "it may not be the answers that are wrong; we may simply be asking the wrong question."

Nevertheless, Lucas (2003, p. 11) considers the effect of making monetary policy more predictable and concludes that the potential welfare gain is indeed small, "on the order of hundredths of a percent of consumption."

In an introduction to his collected papers published in 2013, writes that his conclusion in the 2003 address was that in the postwar era in the U.S., monetary factors had not been "a major source of real instability over this period, not that they *could* not be important or that they never had been. I shared Friedman and Schwartz's views on the contraction phase, 1929-1933, of the Great Depression, and this is also the way I now see the post-Lehman 2008-2009 phase of the current recession." (Lucas 2013, p. xxiv, italics in the original.) In effect, he retreats and concedes Cochrane's point, that it would have been more important to study the role of the Fed as lender of last resort.

Lucas (2003) also goes out on a limb by endorsing Prescott's (1986) calculation that 84% of output variability is due to phlogiston/technology shocks, even though Cochrane also reported results showing that the t-statistic on this estimate was roughly 1.2, so the usual two standard-error confidence interval includes the entire range of

possible values, $[0\%, 100\%]$. In fact, Cochrane reports that economists who tried to calculate this fraction using other methods came up with estimates that fill the entire range from Prescott estimate of about 80% down to 0.003%, 0.002% and 0%.

The only possible explanation I can see for the strong claims that Lucas makes in his 2003 lecture relative to what he wrote before and after is that in the lecture, he was doing his best to support his friend Prescott.

7.2 Example 2: Sargent Supports Lucas

A second example of arguments that go above and beyond the call of science is the defense that Tom Sargent offered for a paper by Lucas (1980) on the quantity theory of money. In the 1980 paper, Lucas estimated a demand for nominal money and found that it was proportional to the price level, as the quantity theory predicts. He found a way to filter the data to get the quantity theory result in the specific sample of U.S. data that he considered (1953-1977) and implicitly seems to have concluded that whatever identifying assumptions were built into his filter must have been correct because the results that came out supported the quantity theory. Whiteman (1984) shows how to work out explicitly what those identifying assumptions were for the filter Lucas used.

Sargent and Surico (2011) revisit Lucas's approach and show when it is applied to data after the Volcker deflation, his method yields a very different result. They show that the change could arise from a change in the money supply process.

In making this point, they go out of their way to portray Lucas's 1980 paper in the most favorable terms. Lucas wrote that his results may be of interest as "a measure of the extent to which the inflation and interest rate experience of the postwar period can be understood in terms of purely classical, monetary forces" (Lucas 1980, p. 1005). Sargent and Surico give this sentence the implausible interpretation that "Lucas's purpose ... was precisely to show that his result depends for its survival on the maintenance of the money supply process in place during the 1953-1977 period" (p. 110)

They also misrepresent the meaning of the comment Lucas makes that there are conditions in which the quantity theory might break down. From the context it is clear that what Lucas means is that the quantity theory will not hold for the high-frequency variation that his filtering method removes. It is not, as Sargent and Surico suggest, a warning that the filtering method will yield different results if the Fed were to adopt a different money supply rule.

The simplest way to describe their result is to say that using Lucas's estimator, the exponent on the price level in the demand for money is identified (in the sense that it yields a consistent estimator for the true parameter) only under restrictive assumptions about the money supply. Sargent and Surico do not describe their results

this way. In fact, they never mention identification, even though they estimate their own structural DSGE model so they can carry out their policy experiment and ask "What happens if the money supply rule changes?" They say that they relied on a Bayesian estimation procedure and as usual, several of the parameters have tight priors that yield posteriors that are very similar.

Had a traditional Keynesian written the 1980 paper and offered the estimated demand curve for money as an equation that could be added into a 1970 vintage multi-equation Keynesian model, I expect that Sargent would have responded with much greater clarity. In particular, I doubt that in this alternative scenario, he would have offered the evasive response in footnote 2 to the question that someone (perhaps a referee) posed about identification:

Furthermore, DSGE models like the one we are using were intentionally designed as devices to use the cross-equation restrictions emerging from rational expectations models in the manner advocated by Lucas (1972) and Sargent (1971), to interpret how regressions involving inflation would depend on monetary and fiscal policy rules. We think that we are using our structural model in one of the ways its designers intended (p. 110).

When Lucas and Sargent (1979, p. 52) wrote "The problem of identifying a structural model from a collection of economic time series is one that must be solved by anyone who claims the ability to give quantitative economic advice," their use of the word anyone means that no one gets a pass that lets them refuse to answer a question about identification. No one gets to say that "we know what we are doing."

8 Back to Square One

I agree with the harsh judgment by Lucas and Sargent (1979) that the large Keynesian macro models of the day relied on identifying assumptions that were not credible. The situation now is worse. Macro models make assumptions that are no more credible and far more opaque.

I also agree with the harsh judgment that Lucas and Sargent made about the predictions of those Keynesian models, the prediction that an increase in the inflation rate would cause a reduction in the unemployment rate. Lucas (2003) makes an assertion of fact that failed more dramatically:

My thesis in this lecture is that macroeconomics in this original sense has succeeded: Its central problem of depression prevention has been solved, for all practical purposes, and has in fact been solved for many decades. (p. 1)

Using the worldwide loss of output as a metric, the financial crisis of 2008-9 shows that Lucas's prediction is far more serious failure than the prediction that the Keynesian models got wrong.

So what Lucas and Sargent wrote of Keynesian macro models applies with full force to post-real macro models and the program that generated them:

That these predictions were wildly incorrect, and that the doctrine on which they were based is fundamentally flawed, are now simple matters of fact ...

... the task that faces contemporary students of the business cycle is that of sorting through the wreckage ...(Lucas and Sargent, 1979, p. 49)

9 A Meta-Model of Me

In the distribution of commentary about the state of macroeconomics, my pessimistic assessment of regression into pseudoscience lies in the extreme lower tail. Most of the commentary acknowledges room for improvement, but celebrates steady progress, at least as measured by a post-real metric that values more sophisticated tools. A natural question meta-question to ask is why there are so few other voices saying what I say and whether my assessment is an outlier that should be dismissed.

A model that explains why I make different choices should trace them back to different preferences, different prices, or different information. Others see the same papers and have participated in the same discussions, so we can dismiss asymmetric information.

In a first-pass analysis, it seems reasonable to assume that all economists have the same preferences. We all take satisfaction from the professionalism of doing our job well. Doing the job well means disagreeing openly when someone makes an assertion that seems wrong.

When the person who says something that seems wrong is a revered leader of a group with the characteristics Smolin lists, there is a price associated with open disagreement. This price is lower for me because I am no longer an academic. I am a practitioner, by which I mean that I want to put useful knowledge to work. I care little about whether I ever publish again in leading economics journals or receive any professional honor because neither will be of much help to me in achieving my goals. As a result, the standard threats that members of a group with Smolin's characteristics can make do not apply.

9.1 The Norms of Science

Some of the economists who agree about the state of macro in private conversations will not say so in public. This is consistent with the explanation based on different prices. Yet some of them also discourage me from disagreeing openly, which calls for some other explanation.

They may feel that they will pay a price too if they have to witness the unpleasant reaction that criticism of a revered leader provokes. There is no question that the emotions are intense. After I criticized a paper by Lucas, I had a chance encounter with someone who was so angry that at first he could not speak. Eventually, he told me, "You are killing Bob."

But my sense is that the problem goes even deeper than avoidance. Several economists I know seem to have assimilated a norm that the post-real macroeconomists actively promote – that it is an extremely serious violation of some honor code for anyone to criticize openly a revered authority figure – and that neither facts that are false, nor predictions that are wrong, nor models that make no sense matter enough to worry about.

A norm that places an authority above criticism helps people cooperate as members of a belief field that pursues political, moral, or religious objectives. As Jonathan Haidt (2012) observes, this type of norm had survival value because it helped members of one group mount a coordinated defense when they were attacked by another group. It is supported by two innate moral senses, one that encourages us to defer to authority, another which compels self-sacrifice to defend the purity of the sacred.

Science, and all the other research fields spawned by the enlightenment, survive by "turning the dial to zero" on these innate moral senses. Members cultivate the conviction that nothing is sacred and that authority should always be challenged. In this sense, Voltaire is more important to the intellectual foundation of the research fields of the enlightenment than Descartes or Newton.

By rejecting any reliance on central authority, the members of a research field can coordinate their independent efforts only by maintaining an unwavering commitment to the pursuit of truth, established imperfectly, via the rough consensus that emerges from many independent assessments of publicly disclosed facts and logic; assessments that are made by people who honor clearly stated disagreement, who accept their own fallibility, and relish the chance to subvert any claim of authority, not to mention any claim of infallibility.

Even when it works well, science is not perfect. Nothing that involves people ever is. Scientists commit to the pursuit of truth even though they realize that absolute truth is never revealed. All they can hope for is a consensus that establishes the truth of an assertion in the same loose sense that the stock market establishes the value

of a firm. It can go astray, perhaps for long stretches of time. But eventually, it is yanked back to reality by insurgents who are free to challenge the consensus and supporters of the consensus who still think that getting the facts right matters.

Despite its evident flaws, science has been remarkably good at producing useful knowledge. It is also a uniquely benign way to coordinate the beliefs of large numbers of people, the only one that has ever established a consensus that extends to millions or billions without the use of coercion.

10 The Trouble Ahead For All of Economics

Some economists counter my concerns by saying that post-real macroeconomics is a backwater that can safely be ignored; after all, "how many economists really believe that extremely tight monetary policy will have zero effect on real output?"

To me, this reveals a disturbing blind spot. The trouble is not so much that macroeconomists say things that are inconsistent with the facts. The real trouble is that other economists do not care that the macroeconomists do not care about the facts. An indifferent tolerance of obvious error is even more corrosive to science than committed advocacy of error.

It is sad to recognize that economists who made such important scientific contributions in the early stages of their careers followed a trajectory that took them away from science. It is painful to say this so when they are people I know and like and when so many other people that I know and like idolize these leaders.

But science and the spirit of the enlightenment are the most important human accomplishments. They matter more than the feelings of any of us.

You may not share my commitment to science, but ask yourself this: Would you want your child to be treated by a doctor who is more committed to his friend the anti-vaxer and his other friend the homeopath than to medical science? If not, why should you expect that people who want answers will keep paying attention to economists after they learn that we are more committed to friends than facts.

Many people seem to admire E. M. Forster's assertion that his friends were more important to him than his country. To me it would have been more admirable if he had written, "If I have to choose between betraying science and betraying a friend, I hope I should have the guts to betray my friend."

References

- Abramovitz, M. (1965). Resource and Output Trends in the United States Since 1870. *Resource and Output Trends in the United States Since 1870*, NBER, 1-23.
- Ball, L., & Mankiw, G. (1994). A Sticky-Price Manifesto. *Carnegie Rochester Conference Series on Public Policy*, 41, 127-151.
- Baumeister, C., & Hamilton, J. (2015). Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information. *Econometrica*, 83, 1963-1999.
- Blanchard, O. (2016). Do DSGE Models Have a Future? Peterson Institute of International Economics, PB 16-11.
- Bunge, M. (1984). What is Pseudoscience? *The Skeptical Inquirer*, 9, 36-46.
- Canova, F., & Sala, L. (2009). Back to Square One. *Journal of Monetary Economics*, 56, 431-449.
- Cochrane, J. (1994). Shocks. *Carnegie Rochester Conference Series on Public Policy*, 41, 295-364.
- Chong, A. La Porta, R., Lopez-de-Silanes, F., & Shliefer, A. (2014). Letter grading government efficiency. *Journal of the European Economic Association*, 12, 277-299.
- Friedman, M. (1953). *Essays In Positive Economics*, University of Chicago Press.
- Friedman, M., & Schwartz, A. (1963) *A Monetary History of the United States, 1867-1960*. Princeton University Press.
- Hatanaka, M. (1975). On the Global Identification Problem of the Dynamic Simultaneous Equation Model, *International Economic Review*, 16, 138-148.
- Iskrev, N. (2010). Local Identification in DSGE Models. *Journal of Monetary Economics*, 57, 189-202.
- Kydland, F., & Prescott, E. (1982). Time to build and aggregate fluctuations. *Econometrica*, 50, 1345-1370.
- Komunjer, I., & Ng, S. (2011). Dynamic Identification of Stochastic General Equilibrium Models. *Econometrica*, 76, 1995-2032.
- Linde, J., Smets, F., & Wouters, R. (2016). Challenges for Central Banks' Models, Sveriges Riksbank Research Paper Series, 147.

- Lucas, R. (1972). Econometric Testing of the Natural Rate Hypothesis. In *Econometrics of Price Determination*, e.d. Otto Eckstein, Board of Governors of the Federal Reserve System, 50-59.
- Lucas, R. (1980). Two Illustrations of the Quantity Theory of Money, *American Economic Review*, 70, 1005-10014.
- Lucas, R. (1994). Comments on Ball and Mankiw. *Carnegie Rochester Conference Series on Public Policy*, 41, 153-155.
- Lucas, R. (2003). Macroeconomic Priorities. *American Economic Review*, 93, 1-14.
- Lucas, R., & Sargent, T. (1989). After Keynesian Macroeconomics. *After The Phillips Curve: Persistence of High Inflation and High Unemployment*, Federal Reserve Board of Boston.
- Onatski, A., & Williams, N. (2010). Empirical and Policy Performance of a Forward-Looking Monetary Money. *Journal of Applied Econometrics*, 25, 145-176.
- Prescott, E. (1986). Theory Ahead of Business Cycle Measurement. *Federal Reserve Bank of Minneapolis Quarterly Review*, 10, 9-21.
- Reicher, C. (2015). A Note on the Identification of Dynamic Economic Models with Generalized Shock Processes. *Oxford Bulletin of Economics and Statistics*, 78, 412-423.
- Romer, C., & Romer, D. (1989). Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz, *NBER Macroeconomics Annual*, 4, 121-184.
- Sargent, T. (1971). A Note on the 'Accelerationist' Controversy. *Journal of Money, Credit, and Banking*, 3, 721-725.
- Sargent, T., & Surico, P. (2011). Two Illustrations of the Quantity Theory of Money: Breakdowns and Revivals. *American Economic Review*, Vol 101, 109-128.
- Sims, C (1980). Macroeconomics and Reality. *Econometrica*, 48, 1-48.
- Smets, F., & Wouters, R (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review*, 93, 586-606.
- Smolin, L. (2007). *The Trouble With Physics: The Rise of String Theory, The Fall of a Science, and What Comes Next*, Houghton Mifflin Harcourt.
- Solow, R. (1974). Comment, *Brookings Papers on Economic Activity*, 3, 733.

Whiteman, C. (1984). Lucas on the Quantity Theory: Hypothesis Testing without Theory. *American Economic Review*, 74, 743-749.