How Empirical Evidence Does or Does Not Influence Economic Thinking and Theory*

Harald Uhlig†
University of Chicago, CentER, NBER and CEPR

This revision: March 24, 2010

*This paper has been prepared for the inaugural conference of the Institute for New Economic Thinking, Cambridge UK, April 8-11, 2010.

†Address: Harald Uhlig, Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637, U.S.A, email: huhlig@uchicago.edu

> "As regards the scope of political economy, no question is more important, or in a way more difficult, than its true relation to practical problems. Does it treat of the actual or of the ideal? Is it a positive science concerned exclusively with the investigation of uniformities, or is it an art having for its object the determination of practical rules of action?" *John Neville Keynes, 1890, Chapter 2*

1 Introduction

Neither have I made the fundamental contributions nor do I have the deep wisdom that would entitle me to be part of this otherwise illustrious panel. Indeed, anything that I may say correctly, I have learned from these giants in the field - and everything that I shall say incorrectly should not be said in the first place. Voicing deep thoughts about economics should better be left to others! I can do no more than pretend. But here I am. Thus, without further apologies, let me proceed anyhow.

This paper asks, how empirical evidence does or does not influence economic thinking and theory. In particular, which role do calibration, statistical inference, and structural change play?

I shall tackle these questions moving from general to specific. For the general perspective, I examine the following points of view.

1. Economics is a science.
2. Economics is an art.
3. Economics is a competition.
4. Economics is politics.

I then examine specific cases for illustration and debate.

1. Is there a Phillips curve?
2. Are prices sticky?
3. Does contractionary monetary policy lead to a contraction in output?
4. What causes business cycles?

The general points as well as the specific cases each have their own implication for the central question at hand. Armed with this list of implications, I shall then attempt to draw a summary conclusion and provide an overall answer.

2 What is economics

"Economics is what economists do[1]". Perhaps. But they approach it in four different ways.

2.1 Economics is a science

Economics is a science. As such, it concerns itself with the description and explanation of economic phenomena. "Positive economics is in principle independent of any particular ethical position or normative judgments" (Friedman, 1953). It describes "what is", as opposed to normative economics, which deals with "what ought to be" (Keynes, 1890).

Positive economics can be used to derive normative implications, though. Economics seeks to find allocations which make humans as happy as possible, taking preferences as given, i.e., taking as given the individual comparison of circumstances as to how happy it will make them. Seeking to improve on the human character, i.e. changing her or his preferences, or dictating ones own preferences as valid for someone else, is typically not only outside the scope of economics, but is generally frowned upon by economist as an approach to crafting economic policy. Indeed, this conflict is often at the heart of public debates, in which economists take one of the sides. The positive economic theorist often seeks to solve for the Pareto optima in her or his model: if one therefore buys into the premises of what makes people happy, then these Pareto optima are indeed the set of normatively best solutions possible.

Economics as a science deduces empirical predictions from theories and induces theory from empirical observations. Both approaches are of importance and mutually complement each other. Smith (1776) deduced deep insights and predictions from the first principles of economic self-interest. So did Menger (1871), the founding father of the so-called "Austrian School of economics" and developer of the concept of marginal utility. Induction in economics could perhaps be associated originally with the "Historical School",

[1]This phrase has been attributed to Jacob Viner, see Backhouse et al (1997).

led by Gustav von Schmoller (1875), and positing that one cannot trust theories not derived from historical experiences or empirical evidence. Schmoller furthermore argued that there cannot be universally valid economic laws, that cultural specificity is central, that structural breaks are ubiquitous.

This "Methodenstreit der Nationalökonomie", this dispute between Menger and Schmoller finds its echo in the modern debates between the proponents of a theory-led deductive approach and the proponents of an empirically grounded inductive approach, all the way to the extreme positions. There are those which reject empirical or econometric techniques altogether, at the one end. At the other end, there are those who reject structural modeling and argue that only natural experiments are valid sources of information. Both sides tend to argue for their perspective with vigor. There, economics is not different from any other science. And as in any other science, both perspectives contribute to its progress in the end. These debates are fruitful and crucial, when viewed from that angle.

I therefore reach a first, tentative conclusion on the question of whether empirical evidence does or does not influence economic thinking and theory. The answer is: it does and it should in economics as a science — but there is not a single successful approach to do so. Whether a theory carefully constructed on a priori grounds is then compared in a "stylized" way with existing evidence, or whether painstaking empirical research leads to summary conclusions for economic theory to ponder: both approaches are valid, always have been and always will be. Not every single researcher may proceed the whole way: indeed, specialization enhances productivity, as Smith (1776) has taught us. But this is and should be an ongoing process for economics as a science as a whole.

This cross-fertilization of theory and quantitative theory on the one side and empirical research and empirical evidence on the other has led to many important breakthroughs.

2.2 Economics is an art

The inductive-empirical ideal of deriving theoretical principles from careful observations or the deductive-theoretical ideal of deriving falsifiable predictions from a priori hypotheses, an approach formulated by Popper (1934), finds its limits in the scientific practice of economics and in the practicability of the application of these principles. Economists cannot and should not aim at explaining it all or understanding it all. Life is too complicated!

Physicists may dream of a "Theory Of Everything" (TOE) or of a "Grand Unified Theory" (GUT). Upon further examination, one finds these dreams to be remarkably limited in scope. Their dream concerns a unified theory for the various kinds of elementary particles as well as the different forces. Physicists dream of providing a final string theory, unifying Einsteins general relativity theory with the standard theory of quantum forces, elementary particles and quantum chromodynamics. Such a theory will be exciting, no doubt, but it will be of little or no additional help in predicting earthquakes, volcanic eruptions or global weather patterns. It will be hopelessly useless for understanding social phenomena such as the causes of war, the sources of poverty and the limits of free markets. If physicists want to call their particular theory a "Theory of Everything", we shall let them. But most certainly, the label is wholly misleading.

Economists, by contrast, do not claim to search for or aim to provide a "Theory Of Everything" (with some exceptions, including perhaps Keynes (1936)). Instead, they aim at answering specific questions. Is there a a trade-off between inflation and unemployment? How can one stimulate economic activity? What aids long-run economic growth? What alleviates poverty? Economists arm themselves with many theories, studying the interaction between a narrow subset of forces at work to answer the particular question at hand. Or they investigate data and summarize these into empirical observations, aimed at shedding light on the particular issue under investigation. Simplification and reduction to the core of the matter at hand, is the key to success in the scientific community of economists. A theory or an empirical investigation should be as simple as possible, but not too simple!

This isn't just Occam's razor, that the hypothesis with the fewest assumption able to explain a given set of facts must be the correct one. It goes beyond the admonition of Hansen and Sargent (2007) to remember the warning of Arthur Goldberger and Robert E. Lucas, Jr. "to beware of theorists bearing free parameters". It truly is a statement of art. We admire the clarity and reductionism of the later works by Piet Mondrian or the structural simplicity of Andy Warhols painting of 100 cans. Mathematicians seek beauty in their proofs. Physicists introduce beauty to guide them in their selection among candidate GUTs. The more beautiful theory wins. It is the scientific principle of minimalistic beauty.

By intended design, the winning beautiful theory ignores many ugly details of realities. By design, a good theory is false. A good theory is not meant to be "realistic". It is meant to incorporate the key aspects that its author intends her or his audience to focus on as important for understanding key aspects of reality. A good theory is a beautiful theory that replicates a selected set of key facts in a convincing and minimalistic fashion. That connection, that judgement takes place outside that theory, however. It takes place within the community of sceptical scientists. As beauty is in the eye of the beholder, scientists will therefore disagree about which theory is the best.

Economics as an art also creates a challenge for the connection to empirical evidence. It makes little sense to "test" a theory according to all its aspects. If a good theory is false by design, then a theory that cannot be directed, is probably not good enough or the data is yet inconclusive. It is necessary to develop econometric techniques that deal with those aspects of the theory, where the theory is meant to generate empirical implications. Likewise, it calls upon empirical researchers to summarize their findings in key facts that a good – and therefore minimalistic – theory is meant to capture.

I therefore reach a second, tentative conclusion on the question of whether empirical evidence does or does not influence economic thinking and theory. The answer is: it influences economic thinking by guiding theorists to design beautiful, minimalistic theories that connect to a select set of key facts. Much gets lost along the way.

There is no doubt that there is a tension between economics as a science and economics as an art.

2.3 Economics is a competition of ideas.

How, then, do economists decide which direction of research is correct, which line of inquiry is fruitful, which argument is convincing, and which one is not? Here, economics – and perhaps science more generally – turns into a competition of ideas.

Economists wish to convince each other. Economics is rhetoric, as McCloskey (1985) has pointed out. Economists wish to receive each others attention. Attention is given to new and novel ideas, to arguments that have not been raised, to insights that run counter to conventional wisdom. Infants turn their heads and pay attention, when they are surprised. Scientists do the same. Humans want to be entertained.

Existing explanations and theories are sometimes abandoned, when empirical evidence falsifies a crucial hypothesis. Those cases seem rare, though.

More likely, existing theories are abandoned because a newer theory is more convincing. Theories are abandoned because its well of novel insights and therefore its resulting steady stream of exciting new discovery and the resulting publications is drying up. Theories are abandoned because the new is sexy and the old is not. This is a theory of scientific revolution as in Kuhn (1970), but with a caveat: the revolutions happen out of boredom with the old and the promise the new territory holds, its a priori appeal. There may be something wrong with that old theory – but the merits lie elsewhere than in cleaning that up.

As a byproduct, old theories are never really discarded. Economists love to hang on to beliefs once formed and they once thought to have learned as correct. Despite the excitement for new ideas that may turn over conventional wisdom, economists cling to it anyhow. One can never take a first look a second time. To take a fresh new look at the facts is difficult individually, and it may be impossible for the science as a whole, perhaps with the exception of the next generation of researchers. As a result, old ideas often have a remarkable staying power.

I therefore reach a third, tentative conclusion on the question of whether empirical evidence does or does not influence economic thinking and theory. The answer is: new theories predict or explain new facts. Nonetheless, old theories stay around, whether they explain existing new and old facts or not.

2.4 Economics is politics

> The ideas of economists and political philosophers, both when they are right and when they are wrong, are more powerful than is commonly understood. Indeed the world is ruled by little else. Practical men, who believe themselves to be quite exempt from any intellectual influence, are usually the slaves of some defunct economist. *John Maynard Keynes, 1936, chapter 24.*

The competition of ideas within the scientific community is a competition for the next beautiful theory or the next beautiful empirical insight. There is no doubt, however, that economic ideas play a powerful role in economic policy. There is therefore another competition of ideas: those that are powerful and convincing in the arena of practical politics. These competitions are distinct. Economists seeking to make their mark as researchers compete in the former. Economists seeking to make their mark as political advisors –

sometimes established as researchers, sometimes not at all – compete in the latter. As a scientist, one wishes that the latter competition be guided by the principles of scientific economics, by the pursuit of Pareto-optimal allocations and well-reasoned arguments, at least in principle. The practical economists readily point to the challenges and compromises in the detail, that they claim must routinely be made. That proposition is rarely, if ever, tested. I fear it is not true as a principle, but is nonetheless followed as a practice due to the limited competition from the scientifically oriented economists, due to the particular procedures governing the selection of political advisors, due to the pressure of rather coming up with a bad answer quickly than a good answer later and due to the limited political resources devoted to find good answers to questions of economic policy at an acceptable pace.

While economics as a science treasures the new and thought-provoking insight, the new and creative argument, the new and powerful evidence, politics is skeptical of new logical arguments, and stays close to the true and proven, to paths well-trodden, to views formed long ago, to formulations that continue to convince the voters. The frustration of John Maynard Keynes (1936) in the quote above is the same frustration that is felt by many in the economics profession today – with the difference, that the defunct economist which John Maynard Keynes wrote about may now be John Maynard Keynes himself. There is a bittersweet consolation in here. Ideas do win, but they win with a long delay, and are often out of synch with the current state of science.

I therefore reach a forth, tentative conclusion on the question of whether empirical evidence does or does not influence economic thinking and theory. The answer is: economic thinking at the practical level of economic policy is thick-skinned and conservative. It is rarely influenced by fresh economic theory or by fresh empirical evidence, unless it fits well with an agenda that had been established elsewhere already.

3 Specific cases

The description above needs to be illuminated by specific cases. This is the purpose of this section. The aim is to generate insights each time on the key question of the influence of empirical evidence on economic thinking.

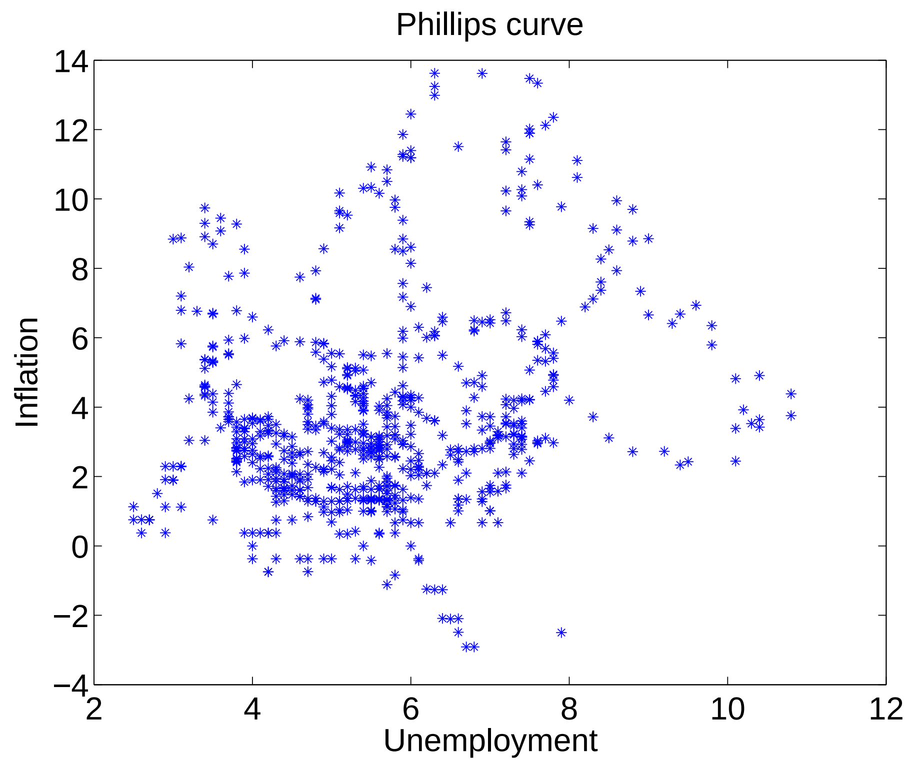

Figure 1: *The US Phillips curve 1948-2008. Data: Civilian Unemployment Rate, BLS, monthly, seasonally adjusted. 12-months log-change in the Consumer Price Index for all urban consumers, BLS, monthly data.*

3.1 Is there a Phillips curve?

The Phillips curve is a tradeoff between unemployment and inflation, originally documented by Phillips (1958) as an empirical relationship in the United Kingdom from 1913 to 1948. It has profoundly influenced economic thinking. Not only is a key ingredient of standard undergraduate textbooks, together with an explanation focussed on sticky wages, it also is deeply embedded in policy debates on monetary policy, for example. As such, one may view this as a poster child for the success of empirical evidence influencing economic thinking.

Well, is it? Figure 1 shows the Phillips curve for the US from 1948 to 2008, juxtaposing monthly unemployment data on the horizontal axis with CPI inflation data on the vertical axis. Figure 2 shows the same data, but

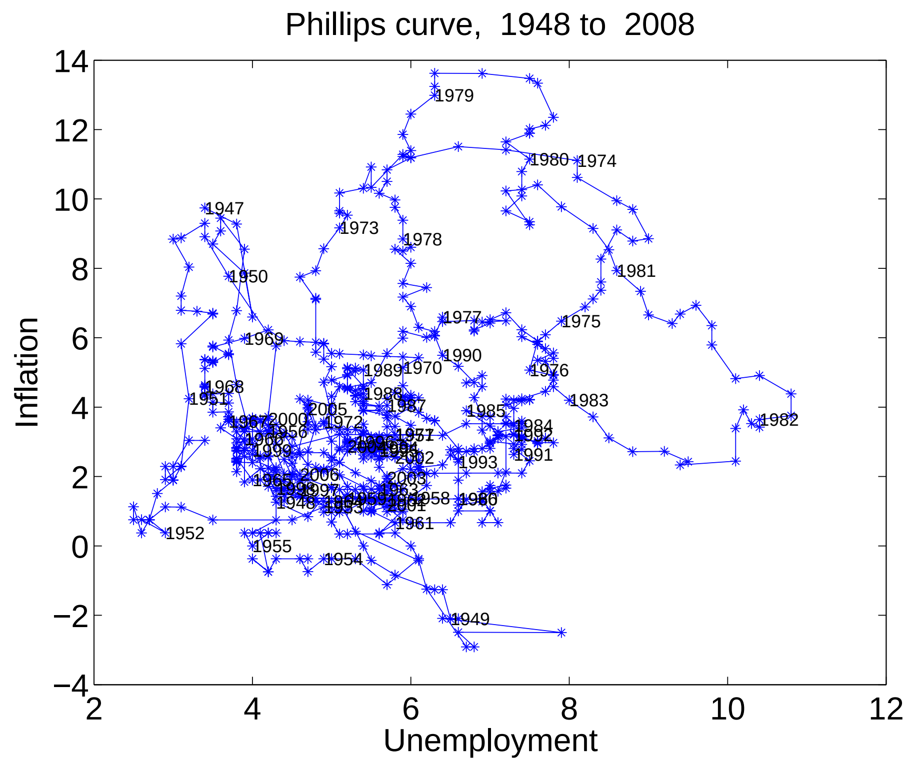

Figure 2: *The US Phillips curve 1948-2008. Adjacent data points have been connected.*

this time "connecting the dots", i.e. connecting adjacent data points with a line, in order to visualize the dynamics. It takes considerable phantasy to see a tradeoff here. Put differently, it is hard to imagine that Phillips would have published this as an interesting relationship, had he seen this in his data.

The panels in figure 3 break the dynamics of figure 2 into segments. There seems to be a fairly stable downward sloping relationship from 1948 to 1959 and certainly from 1959 to 1970. That relationship, however, seems to have broken down from 1970 to 1992, though some may see downward sloping segments here (while blissfully ignoring the upward sloping segments).

The key intellectual insight here is due to Lucas (1976), who has pointed out that the seemingly stable relationship visible in the top two panels of table 3 cannot be expected to be remain stable, once a policy seeks to exploit that relationship. Indeed, the instability from 1970 to 1992 can be seen as a verification of the hypothesis of the rational expectations revolution and the tools of quantitative theory in predicting the outcome of policy experiments. Fed chairmen subsequent to the disinflation episode under Volcker in the early 80's have indeed generally refrained from further attempting to exploit the "Phillips curve tradeoff". The Phillips curve now appears to be nearly flat, judging by the data from 1992 to 2008. The current mainstream view is that the central bank has succeeded in stabilizing inflation, whereas something else is moving unemployment. This perspective offers a new, different perspective on the data, which in turn informs ongoing empirical research. One may view this as a poster child for economic theory influencing empirical insights.

There are two possible conclusions here. Perhaps, there indeed still is a Phillips curve: one just has to look hard enough at these pictures, and explain away episodes that do not quite fit, perhaps with an occasional structural change. Perhaps one needs to focus on surprise innovations or on changes in the inflation rates. Indeed, the conventional current view is that there is still such a tradeoff, but that it cannot be exploited for an extended period of time. Considerable empirical research has gone into re-establishing the Phillips curve tradeoff, but at a more sophisticated level than the simple tradeoff one was originally seeking to see in a figure such as 1. One would need to wonder, though: wouldn't the bottom left panel of table 3 indicate that large changes in unemployment are possible at the small cost of a small change in inflation? Something cannot be quite right here and other forces must be moving both series still.

The other conclusion is that there is in fact no Phillips curve. It is hard to imagine that we would arrive at any other conclusion, if someone were to present figure 1 afresh for the first time today. However, this thinking is too radical.

The Phillips curve has arrived many years ago and it is here to stay. One may be tempted to say: this is so, despite of the empirical evidence accumulated over the last four decades. Do we know what the empirical evidence says? Are we "discovering" the empirical evidence that fits our view of the world that was established long ago? Are we looking at the data with glasses tinted by our theoretical prejudices? Have we been ignoring the empirical evidence all along? Or has the empirical evidence perhaps always been there, as a number of macro-econometricians forcefully argue?

Therefore, on the question of whether empirical evidence does or does not influence economic thinking and theory in this case, one may ask: do we truly know what the empirical evidence says?

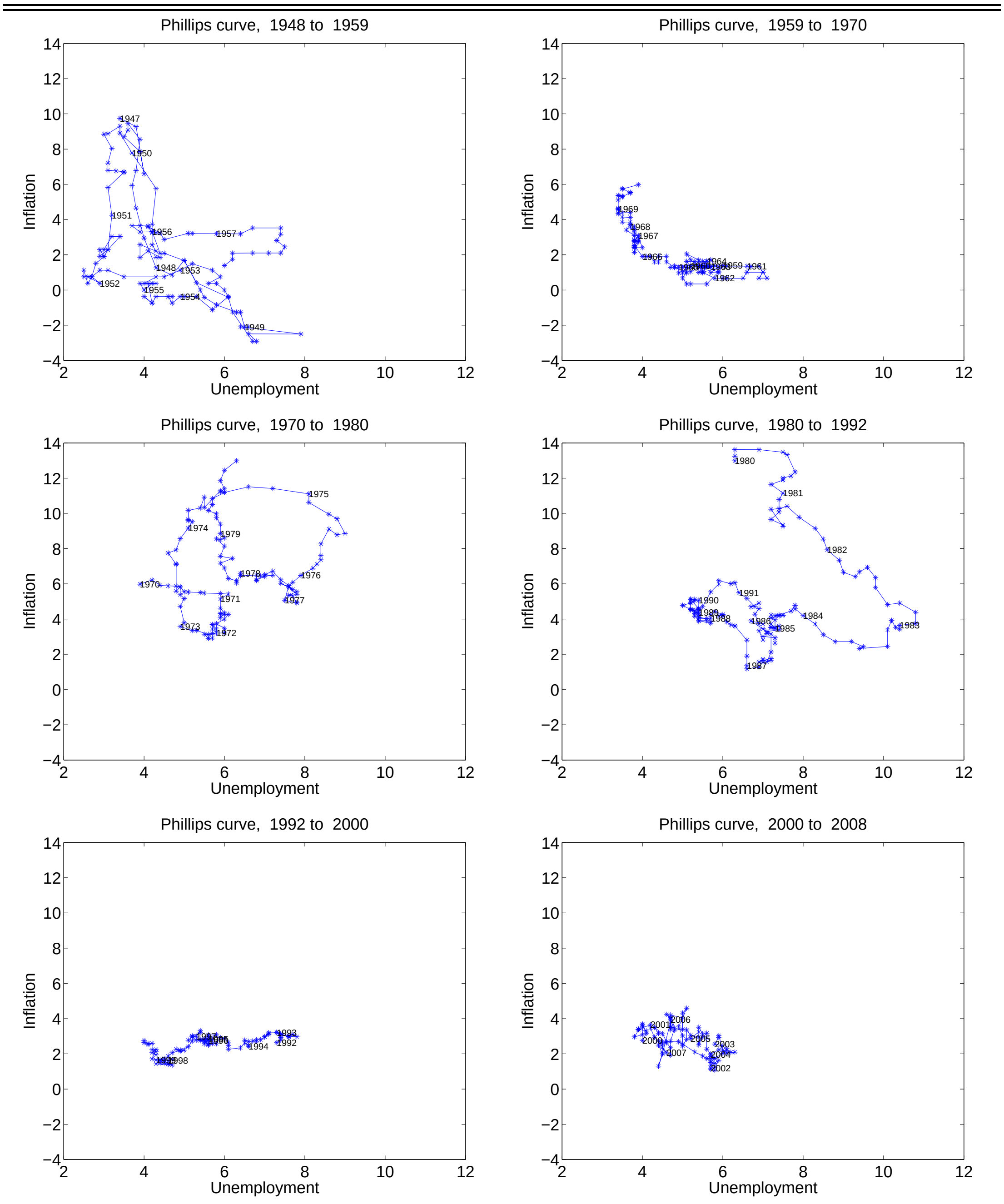

Figure 3: *The changing Phillips curve over time*

3.2 Are prices sticky?

The theoretical explanation for the Phillips curve and its potential exploitability for policy rests on the argument that prices are sticky. Indeed, macroeconomic research in the 80s has often been dismissed as "unrealistic", because it did not feature sticky prices. It seemed self-evident to observers that prices are sticky indeed.

Well, are they? The top graph of figure 4 provides prices for one particular supermarket item - Nabisco Premium Saltines - from one particular supermarket chain, taken from Rotemberg (2005). It may be dangerous to generalize the insight here - indeed, we now know that the general picture is considerably more complex. Nonetheless, there are some important lessons here. The figure shows wholesale and retail prices. On both, one can see "icicles" hanging from some more stable "reference price" level. These icicles represent sales.

Price series of this type have come under great scrutiny in recent empirical macroeconomic research. One possibility here has been to ignore the sales prices, and focus on the reference price instead, see Eichenbaum-Jaimovich-Rebelo (2009). It may be interesting indeed to figure out, why the price returns to the same level after a sale.

Another possibility - and pursued considerably more rarely - is to argue that the price sensitive customers will pay attention to sales and sales prices, and that the interesting economic activity mostly happens at the sales prices. I have therefore connected the sales prices of the top graph by a red line in the middle graph: that is indeed the relevant price series for the dedicated bargain shopper. The reference price is irrelevant for her or him. Removing that irrelevant price data is done in the bottom panel of table 4. It takes a lot of phantasy to see a sticky price series here.

There are many items, for which this description of the empirical facts does not fit. Newspaper magazines in particular have a very stable newsstand price and subscription price, as is well known. However, while the price stays the same, the product changes. Indeed, it changes from one week to the next. Nobody would keep on buying the same old issue, right? Technically speaking, a different product is sold every week, and these products happen to have the same price. The quality of content may change as may the volume of advertisement (a key source of revenue for publishers). Is there evidence here for sticky prices? It is difficult to see how.

So, are prices sticky? Again, it seems hard to say, what exactly the

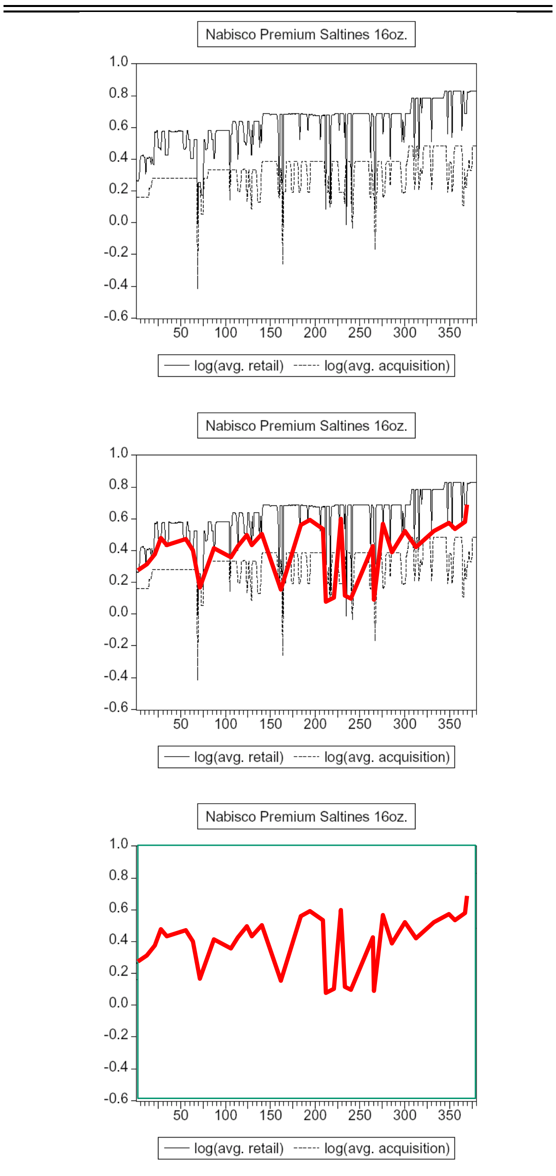

Figure 4: *Nabisco saltines wholesale and retail prices: regular prices and sales prices.*

empirical facts are. If one is to focus on the reference price, or if one is to focus on the constant price for the constantly changing weekly magazine, one may conclude that prices are sticky indeed. But the story is arguably more complicated. The closer one looks, the more one feels that we are asking the wrong question of the data. On the question of whether empirical evidence does or does not influence economic thinking and theory, one feels that theory has yet to ask the interesting questions that the data actually answers.

3.3 Does contractionary monetary policy lead to a contraction in output?

Much of the macro-econometric research in the 60s and the 70s was channeled into constructing large models, incorporating many behavioral equations from the macroeconomic theory of the time, with each equation involving a small number of variables. These equations were then estimated, often assuming the other variables to be exogenous. Given that these equations came from theory, it is hard to say whether the estimation strategy was ever given the chance to "discover" empirical evidence at odds with the theory, and thereby inform and influence economic thinking and theory.

The rectification of these circumstances did not come about by carefully seeking out possibilities for the channel running in reverse from evidence to theory. Rather, a scientific revolution swept these models away (at least in terms of academic interest, as these models are still much in use for policy advise and forecasting). In consequence of the rational expectations revolution, Sims (1980) pointed out, that many theoretical relationships will involve expectations. Since these expectations are formed based on all available present and past data, that consequently means that all present and past data need to appear in all estimation equations. He pioneered the use of vector autoregressions as a tool to handle these relationships: all variables enter as lags in all equations.

The challenge in macroeconomic research then became to sort out the contemporaneous influences of the variables on each other. There, the issue of identification arises even more sharply. If, as a surprise, output falls and interest rates rise, was it the output fall causing interests to rise? Or was it the rise in interest rates causing output to fall? It it is hard and perhaps impossible to answer the question on empirical grounds alone: a theoretical perspective is necessary. One needs to look at the data with glasses tinted by prior views in order to make progress. There is nothing wrong with that. The danger lies in mistaking the general a priori theory tint in the glasses for evidence from the data, which it is not.

Sims (1992) has pointed out, that a priori reasonable strategies for identifying monetary policy shocks, for example, result in odd implications such as the "price puzzle" or the "liquidity puzzle". Put differently, the results seem to be at odds with a priori theorizing. What conclusion ought one to draw? Is the problem the identification strategy? Or is the problem the a priori theory?

Running with the first conclusions, large parts of the literature has tried harder to seek identification procedures, that generate theory-consistent implications. Perhaps indeed, these new identification schemes are the correct ones to use. The new identification schemes may be presented as arising from theory-free a priori reasoning, but it is hard to believe that they weren't found as some implicit specification search and due to the puzzles documented by Sims (1992) in the first place. But if so, then what exactly is now a conclusion and what is an assumption?

Leamer (1978, 1983) has admonished us to take the "con" out of econometrics, and to make such specification searches explicit. One appealing way to proceed in the specific circumstances here is the utilization of sign restrictions, see Uhlig (2005). The central idea is to be clear about the assumptions that are imposed per a priori theorizing. For example, suppose we wish to find out how a surprise rise by the central bank of its key interest rate influences output. One may note that many theories imply that such a monetary policy surprise rise in interest rates leads to a fall in price levels and to a fall in, say, nonborrowed reserves: perhaps not immediately so, but surely with a few months of delay. The theories would rule out that, say, prices or nonborrowed reserves will rise due to a monetary policy surprise rise in interest rates.

These theories may also predict that output will fall. But if that is the focal question, then one surely should not impose that restriction a priori: this is a conclusion we wish to draw from the data, and not an assumption we ought to make a priori, if we want to find out! Only the other sign restrictions ought to be imposed.

Figure 5 shows the results from that exercise: it is documented in detail in Uhlig (2005). While the responses of interest rates, prices and monetary measures are now theory-consistent per construction, there is no clear direction for output. Output may go up or may go down after a surprise that generates the movements shown in all the other variables.

The best interpretation here seems to be that the empirical evidence does not speak loudly on the issue. The theory may be correct that output goes down after a contractionary monetary policy shock. Perhaps, these monetary policy shocks are rare. Perhaps, shocks other than monetary policy generate the pattern in all other variables plus increases in output: superimposed with the output-contractionary monetary policy shocks, this can generate the results in figure 5. The important insight here, however, is this: the data, at least the data used for figure 5 and with the stated a-priori theoretical reasoning employed for identification, does not tell you one way or the other.

The other interpretation is that there is something wrong with the theory. Either e.g. prices ought to rise perhaps temporarily after a contractionary monetary policy shock, as Lawrence Christiano and others have argued. Or output does not show a sharp decline after a contractionary monetary policy shock.

These empirical results have led to some head scratching, but have not truly shifted thinking on what monetary policy surprises do - neither in economic theory nor in monetary policy thinking. The standard view of what a contractionary monetary policy shock is deeply entrenched. Perhaps it is correct. But perhaps it is not. On the question of whether empirical evidence does or does not influence economic thinking and theory, one may be tempted to conclude, that the influence is only there, if it fits the a priori theorizing, while other evidence is dismissed.

At a Carnegie-Rochester conference a few years back, Ben Bernanke presented an empirical paper, in which the conclusions nicely lined up with a priori reasoning about monetary policy. Christopher Sims then asked him, whether he would have presented the results, had they turned out to be at odds instead. His half-joking reply was, that he presumably would not have been invited if that had been so. There indeed is the danger (or is it a valuable principle?) that a priori economic theoretical biases filter the empirical evidence that can be brought to the table in the first place.

3.4 What causes business cycles?

Traditional textbooks used to feature a number of causes for business cycles, resulting in fluctuating labor input into a stable production function, see figure 6. Consider first the concave production function given by $y = g(n)$, with $0 = g(0)$, i.e. no fixed costs. Concavity implies that marginal productivity

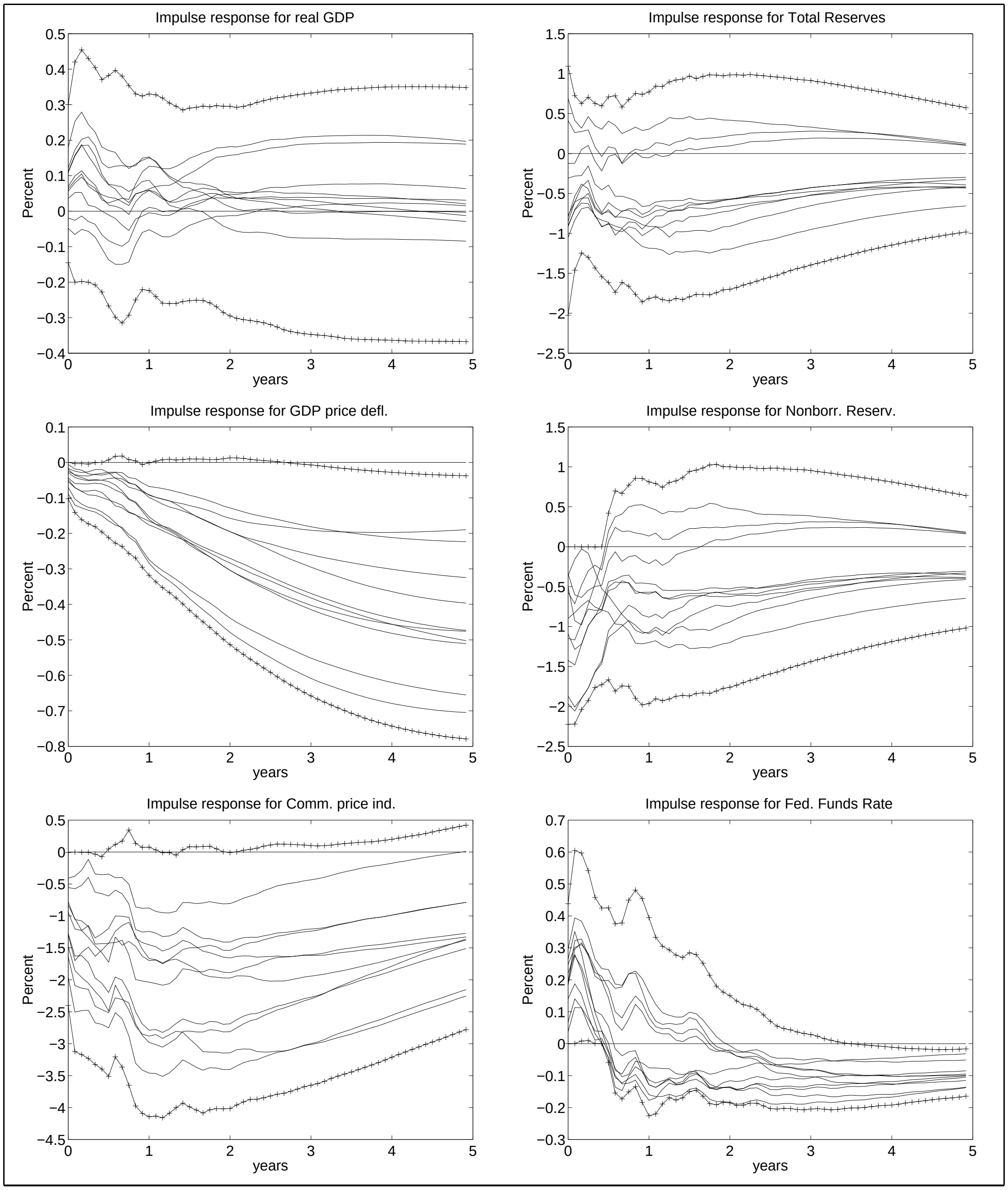

Figure 5: *This figure shows the possible range of impulse response functions when imposing the sign restrictions for $K = 5$ at the OLSE point estimate for the VAR.*

declines with increases in labor input, see the second panel of figure 6. Moreover, as there are no fixed costs, average labor productivity also declines, see the third panel. With this production function and with labor expanding in booms and shrinking in recessions, labor productivity should be high in recessions (point A) and low in booms (point C).

Kydland and Prescott (1982) pointed out that this is at odds with the data. Indeed, as the comparison of the cyclical component of real GDP and labor productivity in figure 7 shows, labor productivity is pro-cyclical, not counter-cyclical. They have pointed to this phenomenon as key to understanding business cycles. Rather than envisioning the production function as stable, they argue that it is is fluctuations in a multiplicative factor for the production function – the technology shocks – which cause this positive correlation and therefore business cycles. Booms are times with high total factor productivity, causing firms to hire lots of labor and a high marginal productivity, whereas recessions are times with low total factor productivity, low levels of labor input and low labor productivity.

The real business cycle theory was key to much of the developments of dynamic macroeconomic theories in the 1980s and 1990s. One may wish to view this as a poster child for the influence of empirics on economic thinking and theory: it was the evidence of pro-cyclical productivity that led to a scrapping of the existing business cycle theories and an entirely new and productive branch of macroeconomic theory. Was it a poster child?

The real business cycle doctrine has been under much attack, for a variety of reasons. The attacks may not have been entirely driven by a fresh, new look at the evidence and an open, unbiased quest for scientific truth. Rather, the doctrine was at odds with existing Keynesian views of macroeconomic dynamics, even though the theories then did not seem to do struggle quantitatively, given the arguments in figure 6 shows. The real business cycle doctrine also was at odds with the prevailing wisdom in economic policy: in its pure form, the theory implies that recessions are strokes of bad luck, but resulting in Pareto-optimal allocations nonetheless. Policy should abstain from countercyclical stabilization measures.

The real business cycle literature and its calibration approach of comparing the theoretical implications with the data was furthermore criticized as lacking sufficient empirical depth, see e.g. Hansen and Heckman (1996). Indeed, many proponents of real business cycle theory redefined the rules of how empirical evidence should influence economic thinking and theory, as far as their applicability to business cycle theory was concerned, see Cooley-

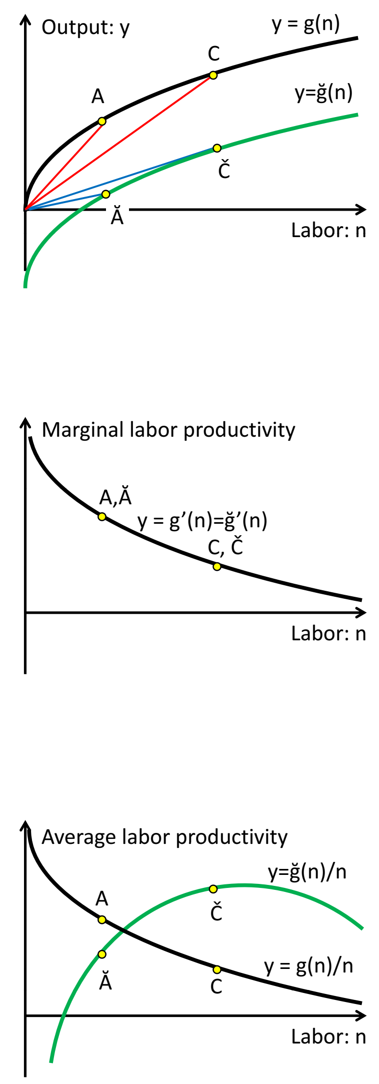

Figure 6: *Production function, marginal and average labor productivity: theory.*

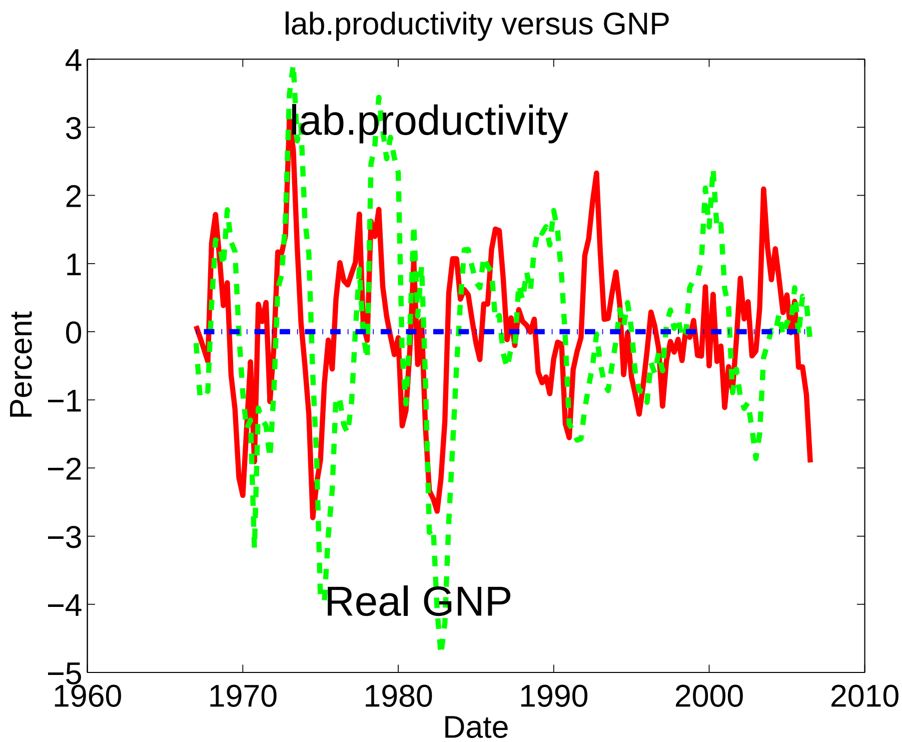

Figure 7: *Comparing the cyclical (HP-filtered) component of labor productivity and real GNP*

Prescott (1995).

During much of the 90s as well as the last decade, new generations of dynamic stochastic general equilibrium models have been developed, responding to these debates. The Smets-Wouters (2003) model has perhaps become the new benchmark for that literature. That model incorporates Keynesian as well as real business cycle features: it could be called a new synthesis. The model features sticky prices, sticky wages and a number of shocks.

Notably, productivity shocks are no longer the key mover of business cycles in Smets-Wouters (2003): they are still there, but of minor importance. To obtain procyclical movements in average labor productivity, Smets-Wouters appeal to fixed costs of production, see $y = \hat{g}(n)$ with $0 > \hat{g}(0)$ in the first panel of figure 6. While a concave production function will still have declining marginal productivity of labor, see the second panel of figure 6, the average productivity of labor may now well be increasing with labor, see the third panel. The model therefore opens standard demand channels as key drivers of business cycles, that are nonetheless in line with the fact that was key to Kydland and Prescott (1982). The Smets-Wouters model and its variants put the route cause of business cycles back to where it was according to Keynes (1936). This is by design of the theory and per views held a priori, rather than by new empirical evidence.

Something else happened to the empirical evidence, though. Aside from debates of how to actually measure total factor productivity, the cyclical correlation between labor productivity and output has shifted since Kydland and Prescott (1982), see figure 7. Has there been a regime shift? Perhaps so. Output fluctuations have become more moderate - and labor productivity has become acyclical. Some attribute the moderation in output fluctuations to changes in the monetary policy rule. If so, however, then the TFP shocks should play a relatively larger role in the second half of the data than the first half: if anything, that should have made labor productivity more procyclical rather than countercyclical, according to typical current theories. There may be a regime shift here, but its implications have not yet been sorted out fully in the macrodynamic theories: once they have, they may contribute to a new re-thinking of the causes of business cycles.

The Smets-Wouters (2003) model was estimated using Bayesian techniques. This serves two purposes. First, it therefore rises to the demand that dynamic stochastic equilibrium models should confront data using established econometric tools, rather than rely on calibration techniques alone. Second, it responds to the defense of the calibration proponents, that an unrestricted estimation is likely to lead to meaningless results, when some knowledge about parameters is already available a priori.

The long list of shocks enables Smets-Wouters (2003) to write down a non-degenerate likelihood function for the parameters of the model, when confronted with a vector of macroeconomic time series. It also has the additional, charming implication, that the model is capable of "explaining" or interpreting all the movements in that data, using the shocks of the model. The way these models are often set up, there is a one-to-one mapping between the one-step ahead surprises in the data and the structural shocks in the model.

The Smets-Wouters model (2003) does not prominently feature a financial sector, except perhaps for a "bond premium shock". There are no banks, there are no credit default swaps, there are no mortgage backed securities. The model is perfectly capable of interpreting the 2007-2009 recession through the lense of the shocks in the model, and generate predictions for monetary policy and its Taylor-rule formulation from that.

One could view this is a resounding success for the the influence of empirical evidence on economic thinking and theory: the theory is now estimated fully, using the latest econometric techniques and therefore informed by the movements in the data. The model builds on major developments in macroeconomics in the last 50 years, both on the new classical side as well as the new Keynesian side. It offers a quantitative interpretation of the data and allows quantitative experimentation with policy shocks, keeping fundamental parameters fixed and thereby respecting the Lucas (1976) critique. It generates results that look largely sensible and can usefully be presented and discussed at the highest level of economic policy. It is the synthesis and culmination of decades of efforts in macroeconomic research and the development of estimation techniques. It is what we always wanted!

Or is it? There are many who believe that financial markets and their disruptions are key to understanding the deep recession of 2007-2009. Models that interpret that recessions without prominently featuring such elements are now routinely viewed as a priori suspect. Elements of the model are furthermore criticized as too tailor-made to fit particular features or views. The partial assumptions about the indexation of otherwise sticky prices, for example, may have murky micro-foundation as do adjustment costs to the level of investment. An agenda of examining and correcting these deficiencies is fruitful and is currently pursued, in particular by a number of central bank researchers. It is easy to criticize this and related models, but in science, you either put up or you shut up. Only a better model beats an existing one.

Unfortunately, as a research agenda, it also may be a well running dry quickly in terms of generating possibilities for the next generation of young scientists to make their mark. Rather than pushing this agenda to fruition, new fields of inquiry in macroeconomics are opened up. Are rare and large shocks responsible for equity markets, trade flows and business cycles? Is noise and information heterogeneity a key force for explaining aggregate fluctuations? Are agents inattentive to minor policy changes? What explains the movements in the housing market and what are its aggregate implications? What are good models - i.e. simple, stylized models - of the financial sector and its implications for monetary policy? What role does agent heterogeneity play for macroeconomic dynamics? What is key to understand aggregate labor markets? What are efficient ways for collecting taxes, when respecting incentives for work and for revealing individual productivity? Are robust decision rules and doubts about long-run growth important for understanding asset prices in a macroeconomic context?

These are among the questions currently pursued by the next generation of macroeconomic researchers. The nearly successful agenda of constructing reliable DSGE models may be in the process of being abandoned in the halls of academia (though still pursued in institutions close to economic policy). Just when calibration had been replaced by statistical inference, and when questions about structural change were raised in a quantitatively sophisticated and interesting manner, events destroy the consensus on which these models were built and economic science turns it attention into a different direction.

4 Conclusions

Let me return to the main question guiding this paper. How does empirical evidence influence economic thinking and theory - or how does it not? Which role do calibration, statistical inference, and structural change play? Several answers emerged:

1. It does and it should in economics as a science — but there is not a single successful approach to do so.

2. It influences economic thinking by guiding theorists to design beautiful, minimalistic theories that connect to a select set of key facts. Much

gets lost along the way.

3. New theories predict or explain new facts. Nonetheless, old theories stay around, whether they explain existing new and old facts or not.

4. Economic thinking at the practical level of economic policy is thick-skinned and conservative. It is rarely influenced by fresh economic theory or by fresh empirical evidence, unless it fits well with an agenda that had been established elsewhere already.

In terms of specific cases, additional aspects emerged:

1. Do we truly know what the empirical evidence says?

2. Theory has yet to ask the interesting questions that the data actually answers.

3. One may be tempted to conclude, that the influence is only there, if it fits the a priori theorizing, while other evidence is dismissed.

4. Just when calibration had been replaced by statistical inference, and when questions about structural change were raised in a quantitatively sophisticated and interesting manner, events destroy the consensus on which these models were built and economic science turns it attention into a different direction.

The summary is this. Empirical evidence influences economic thinking and theory and vice versa - but it does not do so in textbook fashion. Jolted by new empirical and theoretical insights and subjected to the fickleness of attention, the frontier of our science lurches forward to the unknown territory of ever more profound understanding. If it moves in circles, it hopefully does so on ever higher levels. Practical economics and economic policy follows, with considerable distance. Perhaps, this is how it has to be.

References

[1] Backhouse, Roger, Roger Middleton and Keith Tribe (1997)," 'Economics is what Economists do', but what do the numbers tell us?", draft, University of Birmingham.

[2] Cooley, Thomas F. and Edward C. Prescott, "Economic Growth and Business Cycles," chapter 1 in T. F. Cooley, ed., Frontiers of Business Cycle Research, (Princeton University Press, 1995), 1-38.

[3] Eichenbaum, Martin, Nir Jaimovich and Sergio Rebelo (2009), "Reference Prices and Nominal Rigidities," NBER Working Paper 1382.

[4] Hansen, Lars Peter and James J. Heckman (1996), "The Empirical Foundations of Calibration," Journal of Economic Perspectives, American Economic Association, vol. 10(1), pages 87-104.

[5] Hansen, Lars Peter and Thomas J. Sargent (2007), *Robustness*, Princeton University Press.

[6] Friedman, Milton (1953), *Essays in Positive Economics*, The University of Chicago Press.

[7] Keynes, John Maynard (1936), *The General Theory of Employment, Interest and Money*, London: Macmillan (reprinted 2007)

[8] Keynes, John Neville (1890), *The scope and method of political economy,* republished by Batoche Books, Kitchener (1999) per http://socserv.mcmaster.ca/econ/ugcm/3ll3/keynesjn/Scope.pdf

[9] Kuhn, Thomas (1970), *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press.

[10] Kydland, Finn E and Edward C. Prescott (1982), "Time to Build and Aggregate Fluctuations," Econometrica, vol. 50(6), pages 1345-70.

[11] Leamer, Edward E. (1983), "Let's take the Con out of Econometrics," American Economic Review, vol. 73, 31-43.

[12] Leamer, Edward E., (1978), *Specification Searches*, John Wiley & Sons, New York.

[13] Lucas, Robert (1976), "Econometric Policy Evaluation: A Critique", in Brunner, K.; Meltzer, A., The Phillips Curve and Labor Markets, Carnegie-Rochester Conference Series on Public Policy, 1, New York: American Elsevier, pp. 1946.

[14] McCloskey, Deirdre (1985), *The Rhetoric of Economics*, Madison: University of Wisconsin Press.

[15] Menger, Carl (1871), *Grundsätze der Volkswirtschaftslehre*, Wien, Wilhelm Braumüller, downloadable as pdf-file here.

[16] Phillips, A. W. (1958), "The Relationship between Unemployment and the Rate of Change of Money Wages in the United Kingdom 1861-1957," Economica 25 (100): 283299.

[17] Popper, Karl (1934), *Logik der Forschung*, in English (1959), *The Logic of Scientific Discovery*, Routledge, Taylor & Francis.

[18] Rotemberg, Julio J (2005), "Customer Anger at Price Increases, Changes in the Frequency of Price Adjustment and Monetary Policy," Journal of Monetary Economics 52, no. 4, 829-852.

[19] Gustav von Schmoller (1875), *Grundfragen des Rechts und der Volkswirtschaft* , republished Vdm Verlag Dr. Müller (2006).

[20] Smets, Frank and Raf Wouters (2003), "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," Journal of the European Economic Association, MIT Press, vol. 1(5), pages 1123-1175.

[21] Sims, Christopher A. (1980), "Macroeconomics and Reality," Econometrica, vol. 48 (1), 1-48.

[22] Sims, Christopher A. (1992), "Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy," European Economic Review, vol. 36, 1992, 975-1011.

[23] Smith, Adam (1776), *An Inquiry into the Nature and Causes of the Wealth of Nations*, republished by the University Of Chicago Press (1977).

[24] Uhlig, Harald (2005), "What are the effects of monetary policy on output? Results from an agnostic identification procedure", Journal of Monetary Economics 52, 381-419.