DSGE MODELS AND THE LUCAS CRITIQUE

Samuel Hurtado

BANCO DE ESPAÑA

Abstract

Modern DSGE models are microfounded and have deep parameters that should be invariant to changes in economic policy, so in principle they are not subject to the Lucas critique. But the literature has already established that misspecification issues cause parameter instability after policy changes in DSGE models too. This paper will look at the implications of parameter shifts for econometric policy evaluation, to see if policy advice derived from DSGE models would have been fundamentally different from that which the policymakers of the 1970s derived from their reduced-form Phillips curves. The results show drift in most parameters, including those that are supposedly structural (like the share of capital in production, habits, or the elasticity of labor supply to the real wage), and very important shifts in the impulse response functions derived from the real-time estimation of the model. After the expansionary monetary shocks of the early 1970s, a standard DSGE model would have behaved very similarly to an old-style Phillips curve, with marked shifts in parameter values and impulse response functions.

JEL: C11, C32, E32, E60

Contact information:

Samuel Hurtado
Banco de España
C/ Alcalá, 48, 28014, Madrid, España
samuel.hurtado@bde.es
Tel. 914488787

1 Introduction

Lucas (1976) called attention to the fact that reduced form models, while very useful for forecasting, were not suitable for econometric policy evaluation: when a change in economic policy is introduced, agents may react, changing their expectations and behavior, and this will shift the parameters of that reduced form model. For example, a policymaker might estimate a Phillips curve and see a trade-off between output and inflation; but if he tried to exploit this trade-off, agents would notice, they would adjust their behavior to their new expectation of higher future inflation, and the effect on output would be smaller than what the policymaker expected; later on, the policymaker would re-estimate his model with new data, and would find that the trade-off was in fact less advantageous than he initially thought. The solution to this issue is to use structural, microfounded models, with deep parameters: by estimating not the relationship between output and inflation, but the coefficients of the utility function of the consumers and of the production function of the firms, the policymaker could attain a model that is invariant to policy changes, and therefore better suited for econometric policy evaluation.

DSGE models should, in principle, meet these criteria. But, as shown by Cogley and Yagihashi (2010) and by Chang, Kim and Schorfheide (2011), a model which is not correctly specified (e.g. because it uses the wrong specification for price stickiness, or because it doesn't take into account agent heterogeneity) will also display parameter instability issues in response to policy shocks, and may provide inaccurate policy advice. Cogley and Yagihashi (2010) use a model with state-dependent pricing to generate data, simulating a policy shift designed to resemble the change in US monetary policy around the time of the Volcker disinflation; they then use that data to estimate a model with time-dependent pricing a la Calvo (1983) and Yun (1996); they find that due to this model misspecification "private-sector parameters shift across regimes, some by an economically meaningful amount", but also that in terms of welfare the cost of using the wrongly-specified model to derive optimal policy is relatively minor. For their part, Chang, Kim and Schorfheide (2011) generate simulated data using a heterogeneous-agents model in which households have to insure themselves against idiosyncratic income risks (an economy where equilibrium outcomes depend on the cross-sectional distribution of households' wealth and earnings, which in turn depend on the policy regime), and then use that data to estimate a representative-agent model; they find that "if the representative-agent model is estimated with data from the heterogeneous-agents economy under different policy regimes, several important parameters vary considerably", including preference and technology parameters; that "the prediction bias due to imperfect aggregation is substantially larger than the prediction intervals that reflect parameter estimation uncertainty"; and also that "the representative-agent model that abstracts from cross-sectional heterogeneity on the household side can potentially mislead fiscal policy predictions", with important welfare losses caused by deriving policy advice from a misspecified DSGE model.

In another branch of the literature, empirical studies have found parameter drift in estimated DSGE models, using very different approaches and methodologies. Fernandez-Villaverde and Rubio-Ramirez (2007) estimate a relatively standard DSGE model, expanded to allow for parameter drift, and find that all the parameters related to the reduced-form mechanisms introduced in the model as a way to match the observed data (Calvo pricing, indexation, rigidities in consumption and investment, Taylor rule, etc) seem to be evolving with time. Their methodology is uniquely elegant, in the sense that agents in the model are aware that these coefficients are changing, and that they may also change in the future[1], but it is also somewhat limited: they don't derive a formal test to check if a parameter is stable or not, and, because of computational requirements (related to the particle filter they have to use for the estimation), they only allow one parameter to change in each exercise, while all the others remain constant, which could lead to identification issues[2]. For their part, Inoue and Rossi (2008) derive a formal test that checks whether there's a break point in the sample, and identifies the subsets of stable and unstable parameters in the estimation of the model. They apply this methodology to the 1959-2004 period, and find that "time variation afflicts not only the parameters in the monetary policy reaction function, but also most of the parameters in the Euler and IS equations".

The first branch of the literature (Cogley and Yagihashi (2010), Chang, Kim and Schorfheide (2011)) looks at theoretical issues (parameter consistency and pseudo-true values) and uses model-generated data, whereas the second branch (Fernandez-Villaverde and Rubio-Ramirez (2007) and Inoue and Rossi (2008)) looks at parameter instability using real-world data but hasn't yet analyzed in detail its implications for policy advice. I intend to bridge that gap in this paper, at least partially.

In section 2, I will illustrate the parameter drift issue, with a rolling-window estimation of the Smets and Wouters (2007) model which allows all the estimated coefficients to evolve over time[3]. In section 3 (which is the main contribution of this paper), I will consider the implications of observed parameter drift for policymaking, looking at the impulse response functions of the estimated model and analyzing how they change over time. I will carry out a pseudo real time exercise, which will wonder if policy advice derived from estimated DSGE models would have been fundamentally different from that which the policymakers of the 1970s derived from their reduced-form Phillips curves. The results from this pseudo real time experiment show very important shifts in the impulse response functions derived from the successive estimations of the model, around the time where expansionary monetary policy was implemented. This resembles very closely one of the examples of Lucas (1976): if your model is subject to the Lucas critique, it may make you wrongly believe that there's a sizeable trade-off between output and inflation that can be exploited, but when you try to do so, the parameters of your model change, and then you find that the re-estimated models show a not-so-advantageous trade-off. Section 4 will conclude.

---

[1] In most other models, agents think that all parameters are fixed and invariant, and that they will remain so for the foreseeable future: even if the exercise shows that all parameters are found to be changing, agents in the model will never be aware of this fact, and they won't consider in their decision making the possibility of future changes.

[2] This issues would be similar to those created by omitted variables in standard regressions: if there's one parameter drifting (for example, the Calvo parameter for wages), but the estimation exercise leaves it fixed and only allows another parameter to drift (for example, wage indexation), it could wrongly conclude that this second parameter is drifting too (if the effects of increasing one or the other are, to some extent, similar).

[3] This is just a visual illustration, section 2 doesn't aim at establishing proof of parameter instability in the data: that has already been done with a lot more scientific rigor by Inoue and Rossi (2008).

2 Rolling window estimations and parameter drift

Fernandez-Villaverde and Rubio-Ramirez (2007) and Inoue and Rossi (2008) present evidence that the parameters of an estimated DSGE model have not been constant over the last few decades. I will use a rolling window estimation exercise to further illustrate this issue. What I will present in this section is only suggestive evidence that supposedly deep parameters may be drifting, not a formal test that they are doing so at any given statistical significance level[4]. The model I will use is that of Smets and Wouters (2007, AER). I chose this model because it has been tweaked for estimation, it is very standard in the literature, and it is similar to the ones used by both Fernandez-Villaverde and Rubio-Ramirez (2007) and Inoue and Rossi (2008). The data for the estimation will also be that used by Smets and Wouters (2007): growth of output, consumption and investment; employment; price and wage inflation; and interest rate; the only change is that I updated and expanded their database to include the latest data and a longer sample period: Smets and Wouters (2007) use 1966-2004, I will use 1948-2011. The estimation uses Bayesian methods, as in Smets and Wouters (2007), and with their prior distributions too.

As in Cantore, Ferroni and Leon-Ledesma (2011), I run a rolling window estimation[5] of this structural model, and look at the changes over time of the estimated coefficients. Chart 1 shows the results for a selection of 12 parameters of the model. The graphs include both the mean of the posterior distribution, and a 90% confidence interval constructed with the data from the Metropolis-Hastings algorithm. Table 1 shows the results for all the estimated coefficients of the model: the value from Smets and Wouters (2007), the average of the means from all the 20-year window estimations, and the maximum and minimum from those means. The column labeled “drift” is an indicator that takes the maximum and minimum point estimations for each parameter, and checks whether the corresponding 90% confidence intervals of the posterior distribution have null intersection[6].

---

[4] For such a formal test, go to the aforementioned paper by Inoue and Rossi (2008). Note that their test has the same internal consistency issues as my rolling window and recursive estimations: even if the agents in the model live in a world where parameter changes are possible –and, in fact, observed– they believe that parameters are fixed and won’t change in the future (but actually, according to Quoidbach, Gilbert and Wilson (2013), that is, to some extent, how humans behave). Also, their test looks for a single structural break that affects all parameters at the same point in time; my sample is too long for just one single break point (the early 1970s, 1984, and the late 2000s, are all good candidates for a break point) and I want to consider the possibility of different parameters changing at different points in time; their test could be extended to allow for several break points, but this would further deteriorate its power (and their own Monte Carlo analysis already shows that unrealistically long sample sizes are required for the test to have high power).

[5] The results presented in this section are from a rolling window of 20 years, but I have also repeated the exercise with 10-year windows and 5-year windows, obtaining results that are qualitatively similar, but slightly more volatile (in any case, even for the 5-year windows, clearly less volatile than those from Fernandez-Villaverde and Rubio-Ramirez (2007).

[6] This indicator could be related to a structural change test based on Canova and Ortega (2000), but given that I’m not taking into account many relevant nuances of this particular estimation (unknown break point, correlation between different estimated parameters, etc), I make no claims of test size, power, consistency, etc. The tests presented in Inoue and Rossi (2008) are a more rigorous way to statistically identify which parameters are stable and which ones are drifting.

Chart 1: Mode and confidence intervals of 20-year rolling window estimation

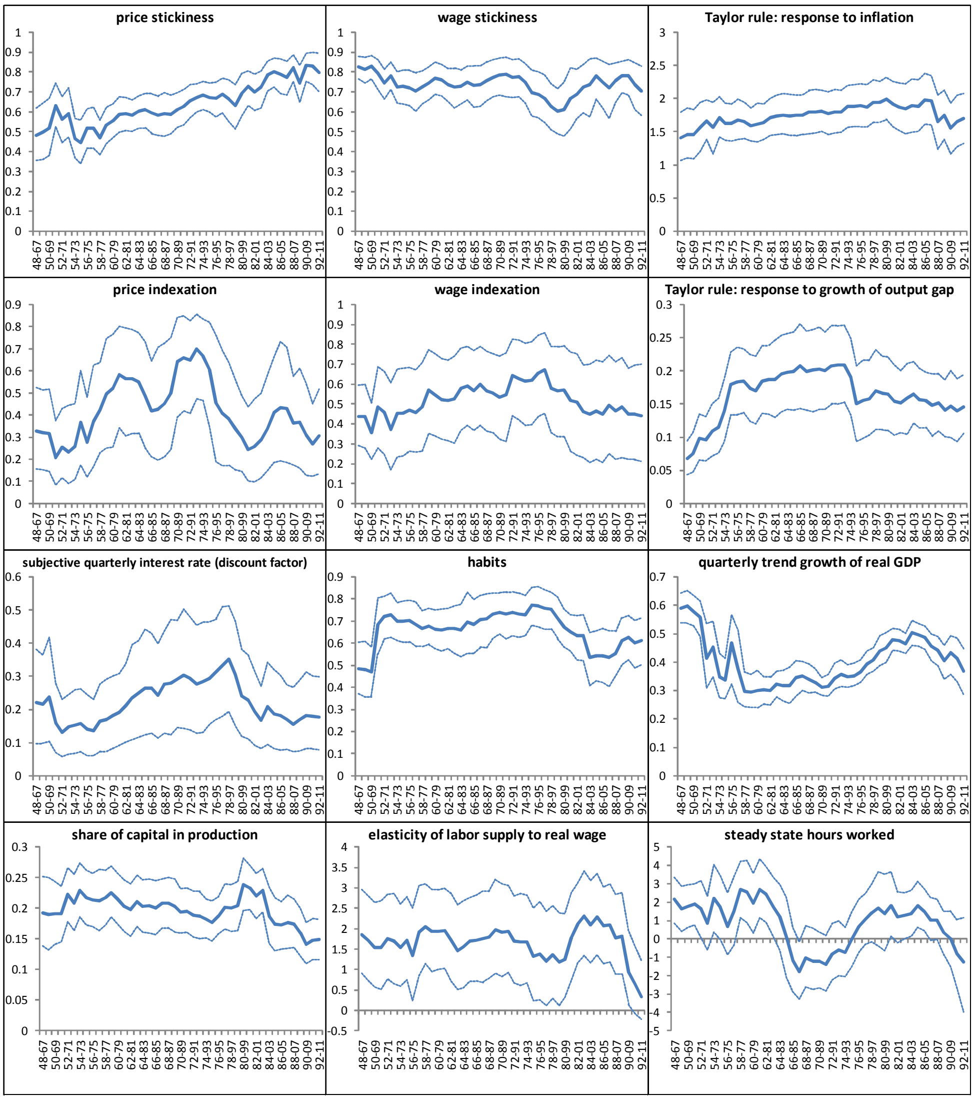


Table 1: results from the 20-year rollling window estimation

	Estimated values				
	S-W	average	min	max	drift
price stickiness	0.65	0.64	0.44	0.83	y
wage stickiness	0.73	0.74	0.60	0.83	y
price indexation	0.22	0.42	0.21	0.70	y
wage indexation	0.59	0.52	0.36	0.67	-
subjective quarterly interest rate (discount factor)	0.16	0.22	0.13	0.35	-
habits	0.71	0.67	0.47	0.77	y
elasticity of intertemporal substitution (for constant labor)	1.39	1.22	0.84	1.68	y
elasticity of labor supply to real wage	1.92	1.66	0.34	2.30	y
share of capital in production	0.19	0.20	0.14	0.24	y
steady state elasticity of the capital adjustment cost function	5.48	4.92	3.54	6.31	-
elasticity of the capital utilization adjustment cost function	0.54	0.50	0.29	0.73	y
fixed cost	1.61	1.52	1.42	1.63	-
Taylor rule: inertia	0.81	0.81	0.71	0.96	y
Taylor rule: response to inflation	2.03	1.75	1.41	2.00	-
Taylor rule: response to output gap	0.08	0.12	0.03	0.18	y
Taylor rule: response to growth of output gap	0.22	0.17	0.07	0.21	y
quarterly trend growth of real GDP	0.43	0.40	0.29	0.60	y
quarterly steady state inflation rate	0.81	0.68	0.56	0.79	-
steady state hours worked	-0.10	0.79	-1.80	2.71	y
stationary TFP shock	0.45	0.47	0.34	0.61	y
price mark-up shock	0.24	0.17	0.10	0.34	y
wage mark-up shock	0.52	0.21	0.14	0.38	y
exogenous spending shock	0.46	0.51	0.36	0.90	y
monetary policy shock	0.24	0.21	0.10	0.34	y
investment spectfic technology shock	0.14	0.47	0.30	0.78	y
shock to interest rate paid/received by households	0.25	0.20	0.06	0.29	y
AR - stationary TFP shock	0.96	0.91	0.77	0.99	y
AR - shock to interest rate paid/received by households	0.18	0.52	0.19	0.95	y
AR - exogenous spending shock	0.98	0.88	0.73	0.98	y
response of exogenous spending to TFP shock	0.52	0.53	0.39	0.64	-
AR - monetary policy shock	0.12	0.27	0.13	0.54	y
AR - investment spectfic technology shock	0.70	0.67	0.36	0.89	y
AR - price mark-up shock	0.91	0.64	0.34	0.92	y
MA - price mark-up shock	0.74	0.55	0.32	0.93	y
AR - wage mark-up shock	0.97	0.81	0.46	0.97	y
MA - wage mark-up shock	0.89	0.61	0.37	0.92	y

The first six coefficients of Chart 1 are those from the Fernandez-Villaverde and Rubio-Ramirez (2007) exercise: Calvo and indexation parameters for prices and wages, and the main parameters of the Taylor rule. These are all reduced-form mechanisms, and the fact that most of these coefficients seem to be evolving over time may not be surprising to many. But the other six should definitely be constant: they are deep, structural parameters, which describe the utility function of consumers, the production function of firms, and the steady state around which the whole model is built.

There are two possible explanations for such drift of supposedly structural parameters: maybe this happens because of misspecification of the model (as illustrated by Cogley and Yagihashi (2010) and by Chang, Kim and Schorfheide (2011)), or maybe it happens because the real world is actually evolving (which, if not captured by the model, is actually a form of misspecification).

If the reason is that the real world is evolving, models will have to take this into account, especially if that evolution is related to economic policy. If things like how much households care about their future, or about their past level of consumption, or about their leisure, are changing over time, they could be doing so in response to different economic policies, and then any model used for econometric policy evaluation should take into account how this process works (how the economic policies it will analyze can alter the behavior of consumers).

If, on the other hand, the behavior of agents in the real world hasn't changed at a deep level (i.e. there are some truly structural parameters that remain constant), and the observed drift in coefficients is due to misspecification of the model, then the source of misspecification should be identified and the model should be "fixed". But it's a daunting task. There are lots of possible sources of misspecification, and finding a DSGE model that can go through the rolling-window estimation exercise[7] without showing parameter drift is not easy[8].

[7] Actually, the proper test would be that of Inoue and Rossi, but its power issues in "small" samples also have to be taken into account.

[8] Part of the problem is that the model has to be estimated: it needs to have the ability to match the observed data, and this is surprisingly uncommon. For example, a near-flexible-prices version of the Smets and Wouters model, with Calvo and indexation parameters set at 0.01, won't lead acceptable estimation values unless priors are narrowed (some other parameters go towards zero or infinity); but narrowing the priors to get a good estimation is not a desirable alternative, because it can be abused: if it is overdone, the stable-parameters result can be forced. Setting those parameters at 0.1 instead of 0.01 makes the estimation work, but is not a solution to model drift anyway: the rolling-window estimation exercise still suggests that several supposedly-deep parameters are evolving over time. The same result is obtained if only the Calvo parameters are set to 0.1, or if Calvo, indexation and habits are all set to 0.1. And same result again for more radical departures: substituting rational expectations (which are convenient and impose discipline but may not be realistic and could be a source of misspecification) for backwards looking expectations doesn't get rid of the drift either, even if they're designed to match the behavior of observed expectations (as proxied by the Survey of Professional Forecasters from the Federal Reserve Bank of Philadelphia).

3 Impulse response functions: the pseudo-real-time exercise

The previous section has shown some indications that supposedly structural models may not be as stable as they should be: deep parameters should remain constant, but a rolling-window estimation of the model shows them evolving over time. These indications are in line with stronger results from previous studies, both for the US economy –Fernández-Villaverde and Rubio-Ramírez (2007), Inoue and Rossi (2008)– and for other advanced economies –Jerger and Röhe (2012).

The biggest problem with this parameter drift is that it could be caused by policy changes, and in that case these models would be unfit to do econometric policy evaluation: a policymaker could use a model to evaluate how the economy would react to his actions, then make his choices and implement some policies, and finally find that the modeled economy has changed and the effect of the policies is not what he anticipated.

In terms of our model, the real problem would be, not that the parameters are drifting, but that the impulse response functions of the model may be changing, and that they may be doing so in response to policy changes: if the coefficients change but the IRF to the shock that represents the policy is mostly unchanged, then parameter drift wouldn't be such a big problem. So it is important to evaluate not only whether parameters are stable, but also how these impulse response functions change when the coefficients of the model evolve over time, especially around changes in economic policy. This could be done with the IRFs derived from the rolling-window estimation presented in the previous section. But instead of that, in this section I will run a pseudo-real-time exercise, looking at the policy advice that a policymaker from the 1970s would have derived from the estimation of the Smets-Wouters model, and comparing that to the performance of an old-style Phillips Curve estimated using ordinary least squares (as a representation of a worst case scenario for the technology available to this hypothetical policymaker).

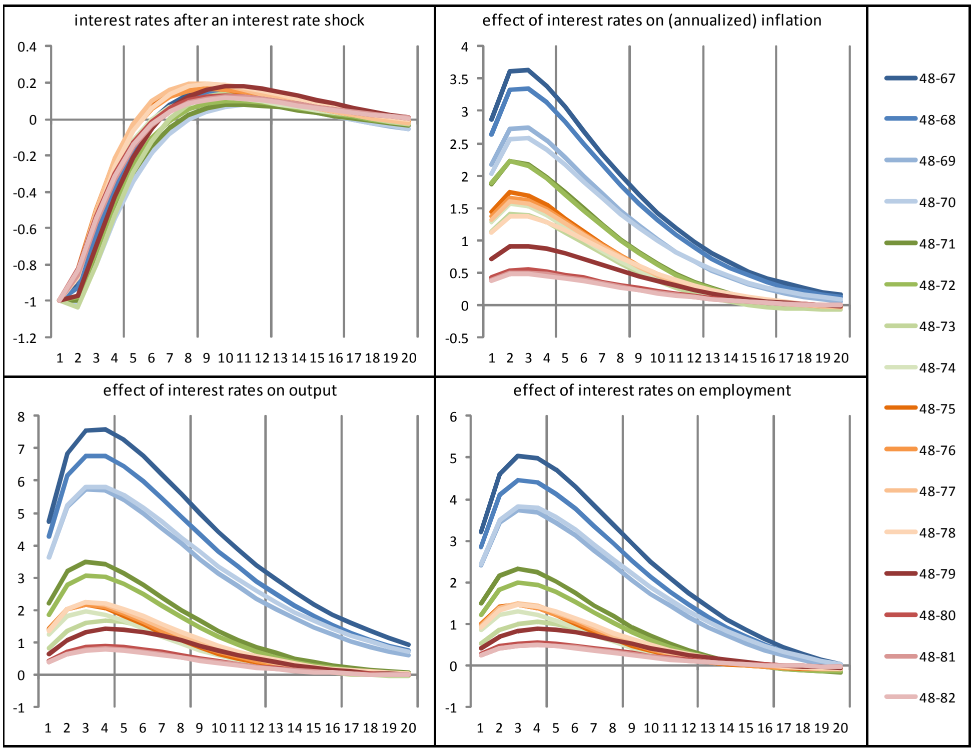
It is often argued[9] that at least part of the economic woes from the 1970s was caused by economic policy mistakes. In particular, that policymakers wrongly identified a trade-off between output and inflation, and tried to exploit it, generating inflation without achieving the expected response of output –and then seeing their estimation of the trade-off becoming a lot less advantageous, once the policies had been applied. Lucas (1976) presents an explanation of why this could have happened: they were estimating Phillips curves, and didn't take into account how their policies would affect the coefficients of those reduced form models. The suggested solution to this problem is to use structural models, with deep parameters whose estimation should remain constant over time, invariant to policy changes. The Smets-Wouters model represents the state-of-the-art evolution of such models, extended and expanded to better fit the data, and similar, for example, to models in use nowadays at many central banks. So in this section I will try to answer the following question: if the policymakers of the 1970s had been using the Smets-Wouters model instead of their old-style Phillips curves, would they have identified the same trade-offs, and would they also have found them to be smaller-than-originally-expected after the policy was implemented?

[9] See, for example Fuhrer et al (2009).

The results from the exercise are summarized in Chart 2, which plots the impulse response functions from each estimated set of parameters, always to a 1 point shock to interest rates, and Chart 3, which takes the responses of output and inflation from Chart 2 and plots the evolution over time of the ratio between them; this is similar to the slope of the Phillips curve, but it has some advantages: the impulse response functions summarize the short-term and long-term trade-offs, and the time frame for each of them; and in practice it is often IRFs and not parameter values which are used to derive policy advice[10].

The exercise is analogous to recursive estimation[11]. Initially, policymakers estimate the Smets-Wouters model with sample period 1948-1967, and look at the impulse response function for a monetary policy shock (a reduction of interest rates of 1 percentage point below the value predicted by the model's Tayor rule). Then, year after year, when new data is available, they extend the end of the sample period (first to 1948-1968, then 1948-1969, etc), re-estimate the model, and look at the new IRF for a monetary policy shock[12].

Chart 2: pseudo-real-time exercise: impulse response functions



10 Additionally, the usual form in which the Phillips curve in DSGE models is presented (as a relation between inflation and marginal cost) is not suitable for this exercise, since it relies in the fact that marginal cost is proportional to the output gap, which is not the case when the parameters of the model are not constant. A form that relates output and inflation can be derived, but the IRF ratio would still be more relevant, since it is closer to what policymakers look at, and it is a better summary of the dynamic tradeoff.

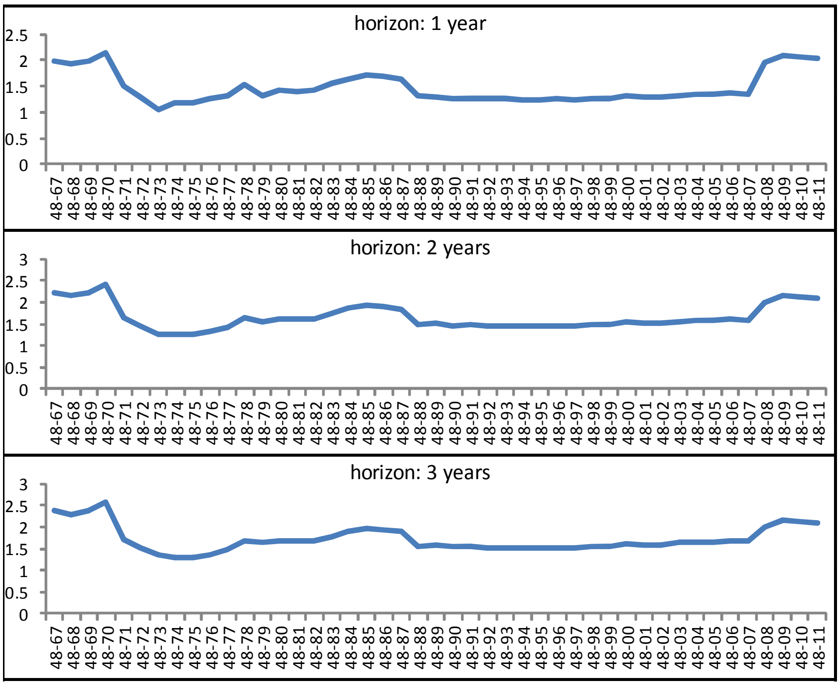
11 This switch from rolling window estimation in the previous section to recursive estimation here is because rolling windows are better at showing the evolution of the parameters over time, but I believe recursive estimation is closer to what central banks actually do. In any case, I have checked that the results would be very similar with a rolling window exercise.

12 The shock is always a 1 percentage point reduction of interest rates on impact, but may be slightly different in the following quarters, depending on the inertia implied by the re-estimation of the model; in any case, differences in this path of interest rates are relatively small, as shown in Chart 2.

The most interesting period is the one around 1970-1972. In the foreword of Fuhrer et all (2009), Paul A. Samuelson cites the very early 1970s as a point of maximum monetary push[13]. And the full-sample estimation of the Smets-Wouters model also identifies expansive monetary-policy shocks in these years. So, this period is a good one to check whether the quantification that the model gives for the effects of a monetary policy shock are constant, or if they're changing as the shocks are implemented in the real world.

As shown in Chart 3[14], the initial estimation (1948-1967) shows a very positive trade-off between output and inflation: a reduction of interest rates generates an increase in output in the first year that is twice as big as the increase in inflation[15]; on average over the first three years, this ratio of the responses of output and inflation is close to 2.5. Then, as time goes by and monetary expansion takes place in the real world, my hypothetical policymakers re-estimate the model year after year, and find that the response of output and employment to this monetary shock is quickly getting smaller. The associated effect on inflation is also getting smaller rather quickly, but by the time the sample period has changed to 1948-1973 (i.e. right after the expansionary monetary shocks were implemented in the real world), the trade-off has already changed dramatically: the ratio of the responses of output and inflation has gone down to one in the first year, and 1.4 on average over the first three years[16].

Chart 3: pseudo-real-time exercise: IRFs: effect on output / effect on inflation



---

[13] In fact, he relates this particular episode to a "Machiavellian purpose of ensuring Nixon's 1972 reelection".

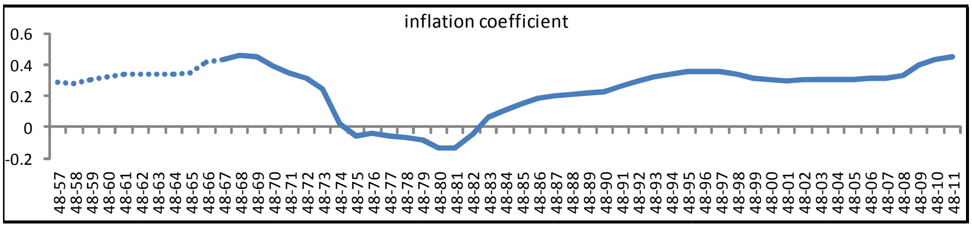
[14] As an example, a value of 2 in the graph for a 3-year horizon means that an interest rate shock that increases the level of output by 2% on average over three years, also increases inflation by one percentage point on average over that same period.

[15] For reference, with the coefficients from Smets and Wouters (2007), who use data from 1966-2004 in the estimation of the model, the response of output is 1.6 times bigger than the response of inflation over the first year, and 1.75 times bigger on average over the first three years.

[16] Whether this change is significative or not may be of little interest: if the policymaker sees that his estimation changes in a noticeable way, he may investigate why that happened, and whether the change is significative or not, but regardless of the results he will use the new impulse response functions for policy evaluation anyway, therefore the shift doesn't need to be significative in order to be relevant.

This is not fundamentally different from what these virtual policymakers would have experienced if they were estimating a simple Phillips curve (with the same specification that is used in the example 5.3 of Lucas (1976): $y_t = a + b\pi_t + \bar{y}_t$) using ordinary least squares. As is shown in Chart 4, the estimated slope of this simple Phillips curve would have evolved in a similar fashion to what we've seen in the IRFs of the Smets-Wouters model, with a marked downwards shift between 1969 and 1975.

Chart 4: pseudo-real-time exercise: simple Phillips curve estimation



This similarity doesn't necessarily mean that the policymakers of the 1970s would have taken the same policy decisions if they were using the Smets-Wouters model to get their policy advice. But it does mean that in terms of parameter invariance and its effect on econometric policy evaluation, the estimated DSGE model wouldn't have done a lot better than the old-style Phillips curve. The lessons that policymakers would have drawn from Smets-Wouters are not fundamentally different from the ones they were extracting from their reduced-form models[17], and a similar experience (believing there's a trade-off that can be exploited, trying to exploit it, and then seeing this policy fail and the estimation of the trade-off shift) seems entirely plausible.

There's no definitive proof here that the shift in the estimated parameters, and in the resulting impulse response functions, is due to the economic policy implemented in those years. Those were convoluted times, and many things were happening all at the same time. But I observe a big change in the evaluation of the effects of a monetary policy right after this policy is implemented, and Cogley and Yagihashi (2010) and Chang, Kim and Schorfheide (2011) show that a misspecified structural model would behave in this way. Also, even if the shift is due to something other than the monetary policy shocks of those years, the simple fact that the impulse response shifts right where the policymakers would want it to be most stable, is a problem; it may not be the textbook case of the Lucas critique (where the policy itself is what causes the shift in the evaluation that the model makes of said policy), but in any case it would be bad news for econometric policy evaluation using DSGE models: something happened that was not contemplated by the model, and which made its evaluation of the policy quantitatively inaccurate.

An additional point of interest could be, which parameters are responsible for this shift in the impulse response functions from the estimated DSGE model. I have tried to identify them with the following procedure: I start with the parameters from the 1948-1970 estimation, and take note of the ratio of IRFs presented in Chart 4; I then start changing one parameter at a time to its value in the 1948-1973 estimation, and see how that ratio of IRFs responds to the change. The results depend on the order in which the parameters are changed from one value to another, but from looking at numerous alternative orderings it is

---

[17] The biggest difference may be that the more modern model imposes that this particular trade-off can only exist in the short run. Given the short-term goal that Paul A. Samuelson attributes to the monetary expansion of the early 1970s, in practice this would probably have made little difference.

safe to say that nearly all of the change comes from the shifts in the Calvo parameter for wages (which goes from 0.84 to 0.75, meaning that average wage duration changes from 6 quarters to 4 quarters), the parameter for habits (which goes from 0.47 to 0.75), the elasticity of labor supply to the real wage (from 1.71 to 2.48), and the elasticity of intertemporal substitution (from 1.01 to 1.23).

4 Conclusion

Previous papers had already shown that the estimation of supposedly deep parameters from supposedly structural DSGE models was far from stable: Fernandez-Villaverde and Rubio-Ramirez (2007), in their very elegant setup, find noticeable drift in the parameters, whereas Inoue and Rossi (2008), with their well crafted statistical tests, find a significant structural change in many of those parameters. More theoretical papers, such as Cogley and Yagihashi (2010) and Chang, Kim and Schorfheide (2011), have shown that something like this may happen because misspecification of the model makes its supposedly-structural parameters non-invariant to policy shocks, sometimes with big consequences for econometric policy evaluation.

This paper, apart from providing further indication of this parameter instability in the data, has looked at how the drift of coefficients in estimated DSGE models would have shifted the policy advice derived from them around the time where these policies were implemented. Since they should be invariant to policy changes, DSGE models were supposed to be the recipe that would provide policymakers with robust advice, but, as shown in the pseudo-real-time exercise of section 3, this parameter drift means that the advice provided by these modern models wouldn't have been fundamentally different from that which the policymakers of the 1970s could derive from their old-style Phillips curves. A model such as Smets-Wouters –similar to the ones in use today in many central banks– would have identified a big short-term trade-off between inflation and output if estimated with data from 1948 to 1970, whereas a re-estimation three years later, after a big monetary expansion occurred, would have shown that the trade-off was in fact much less advantageous. Exactly the problem that Lucas illustrated for reduced-form Phillips curves.

Of course, all of this doesn't mean DSGE models are useless and should be abandoned, but it does indicate that they're not as flawless as many people think they are. For a start, they are not fully structural, since, as Fernandez-Villaverde and Rubio-Ramirez (2007) emphasized, they include several reduced-form mechanisms. But, even beyond that, as shown by Inoue and Rossi (2008), supposedly deep parameters change with time too, maybe in response to changes in economic policy. And of course economic models have advanced a lot over the last 40 years, but the new models, while an improvement over the old ones in many aspects, could eventually lead to similar mistakes when used to evaluate the effects of economic policy, especially in the short and medium term. Using microfounded models is not enough to avoid the Lucas critique, since, as shown by Cogley and Yagihashi (2010) and by Chang, Kim and Schorfheide (2011), and by the pseudo real time exercise in this paper, being a reduced form and being misspecified can have very similar effects in terms of parameter drift after a change in economic policy.

What should be done then? Trying to perfect DSGE models, bringing them closer to reality, finding misspecification issues, and looking for variants that are resilient to the kind of problems presented here, all seem like good ways forward. We should also try to keep their estimation up-to-date –in order to use a version of the model that is as close to recent reality as possible– and watch out for parameter drift. But we should probably also worry a bit less about the risks of adding reduced-form mechanisms or non-rational expectations to these models. A simple mention of the Lucas critique shouldn't be enough to disregard these solutions, since we don't really have –yet– an alternative that is truly immune to that critique.

BIBLIOGRAPHY

Canova, Fabio & Ortega, Eva, 2000. "*Testing Calibrated General Equilibrium Models,*" in Mariano,Roberto & Schuermann,Til & Weeks,Melvyn J. (ed.), 2008. "*Simulation-based Inference in Econometrics,*" Cambridge Books, Cambridge University Press.

Cantore, C. & Ferroni, F. & León-Ledesma, M. A., 2011. "*Interpreting the Hours-Technology time-varying relationship,*" Working papers 351, Banque de France.

Yongsung Chang & Sun-Bin Kim & Frank Schorfheide, 2011. "*Labor-Market Heterogeneity, Aggregation, and the Policy-(In)variance of DSGE Model Parameters,*" RCER Working Papers 566, University of Rochester - Center for Economic Research (RCER).

Timothy Cogley & Takeshi Yagihashi, 2010. "*Are DSGE Approximating Models Invariant to Shifts in Policy?,*" The B.E. Journal of Macroeconomics, Berkeley Electronic Press.

Jesús Fernández-Villaverde & Juan F. Rubio-Ramírez, 2007. "*How Structural Are Structural Parameters?,*" NBER Working Papers 13166.

Jeff Fuhrer & Jane Sneddon Little & Yolanda K. Kodrzycki & Giovanni P. Olivei (ed.), 2009. "*Understanding Inflation and the Implications for Monetary Policy: A Phillips Curve Retrospective,*" MIT Press Books.

Inoue, Atsushi & Rossi, Barbara, 2008. "*Which Structural Parameters Are "Structural"? Identifying the Sources of Instabilities in Economic Models,*" Working Papers 08-02, Duke University, Department of Economics.

Peter N. Ireland, 2007. "*Changes in the Federal Reserve's Inflation Target: Causes and Consequences,*" Journal of Money, Credit and Banking.

Jürgen Jerger & Oke Röhe, 2012. "*Testing for Parameter Stability in DSGE Models. The Cases of France, Germany, Italy, and Spain,*" Working Papers 118, Bavarian Graduate Program in Economics (BGPE).

Jordi Quoidbach, Daniel T. Gilbert, Timothy D. Wilson, 2013. "*The End of History Illusion,*" Science, Vol. 339 no. 6115 pp. 96-98.

Lucas, Robert Jr, 1976. "*Econometric policy evaluation: A critique,*" Carnegie-Rochester Conference Series on Public Policy.

Smets, Frank & Wouters, Rafael, 2007. "*Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach*," CEPR Discussion Papers 6112,