The role of Sticky Information on Inflation Dynamics: Estimates and Findings

Benedetto Molinari*
Universitat Pompeu Fabra

May, 2006

Abstract

I derive and estimate the theoretical second moment of Inflation from Sticky Information Phillips Curve, so to get the degree of Information stickiness in US economy in the last 50 years. The paper makes three points.

First, I test whether the model is accepted by the data. I argue that the estimation strategy I use here is a more powerful test against the alternative model of sticky prices compared with those estimations that exploit the first order moment of Inflation.

Second, the value of Information stickiness I estimate is far away from the standard value used for calibrations (e.g. Makiw and Reis, QJE 2002). This implies that the agents update their information sets more often than what usually assumed, so suggesting that Information stickiness may be only a minor cause of the Inflation persistence observed in the data. In addition, I perform various tests of structural breaks to show how the Information stickiness has changed in the sample. This last evidence is in accordance with the observed change in Inflation persistence during the disinflation of '90, which it is worth notice is an empirical conundrum with the New Keynesian Phillips Curve.

Finally, I show how SI implies endogenous Inflation volatility. In this perspective, the time varying volatility of Inflation observed in US economy, may be due to SI instead of being an evidence of stochastic volatility, as recently argued by Cogley and Sargent (2005), and Canova, Gambetti and Pappa (2005) using TVC-VAR.

JEL Classification: E31, C52

Keywords: Inflation Dynamics, Sticky Information, Time-varying volatility, GMM estimation.

*I'm grateful to Albert Marcet for his guidance and his comments. I wish also to thank Jordi Galí, Kosuke Aoki, and Francesco Turino for useful discussions on this topic. Departament of Economics and Business, Universitat Pompeu Fabra (Barcelona, Spain – EU). Contact Email: benedetto.molinari@upf.edu

1 Introduction

I estimate the Sticky Information (henceforth SI) Phillips Curve using US post-war quarterly data. I get the (most likely) value for Information stickiness in US economy, and I compare it with the standard calibrations used in literature.

There are various reasons why to estimate the SI Phillips Curve. First, although the model belongs to a literature that now raises growing interest, there are few accompanying econometric analyses.

In the last years the research agenda of Inflation and macroeconomic dynamics has bumped into the Rational Inattention hypothesis. This literature takes the move from the early contributions of Lucas (1973), Fischer (1977), Taylor (1980), Woodford (2001), and found a standard modeling in Sims (1998, 2003). Makiw and Reis (2002, henceforth MR) put forward the Sticky Information hypothesis combining some elements of Fisher and Lucas contributions. Their purpose was to explain the gradual response of Inflation to the shocks, as observed in post-war US economy. Their explanation was based on the idea that the firms absorb only sporadically the Information they need for pricing optimally their goods. When the inflow of Information is limited or absent, then the firms are forced to choose suboptimal pricing behavior. The idea is appealing: we could observe a failure of RE not because the firms follow suboptimal behaviors, like indexation or rules of thumb, but because they price (optimally) their goods facing constraints to the Information sets. So, MR model is a first tractable step to investigate the issue of bounded rationality.

Despite the appealing theoretical framework, the empirical analysis of the SI Phillips Curve has been very limited. In the following, I report on the only attempt of estimation I'm aware of,[1] which is highly criticizable though. So, some questions are still pending. MR, and then Trabandt (2003), showed that SI Phillips Curve may be a possible alternative to the New Keynesian Phillips Curve, but do the data actually accept the SI model? To this end, I estimate SI Phillips Curve paying particular attention to choose an econometric strategy which was as powerful as possible test against the alternative hypothesis of New Keynesian Phillips Curve. That's the reason why I discard the estimation based on the first moment specifications of SI model. Not because such estimation is based on some approximations, which are almost unavoidable to estimate this model, but because I'll show it provides a very low power test against the alternative model, which seems to me much more lamentable. In addiction, I analyze which is the actual extent of Information stickiness in the economy. Is it an important issue in the dynamics of Inflation, or it is just a caption remark?

A second reason to estimate SI Phillips Curve is to check whether it can explain two side features of Inflation dynamics that the alternative model, i.e. the New Keynesian Phillips Curve, cannot explain; first, the positive correlation between the level and the persistence of Inflation in the 90's; second, the time varying volatility of Inflation. As I said, SI Phillips Curve was first introduced as a possible alternative to the New Keynesian Phillips Curve, and from the

[1] By Khan and Zhu (2002).

very beginning the two theories has been in competition. However, there are not been found overwhelming evidences in favor of one or the other model, at least regarding the persistence of Inflation. An exercise performed by Trabandt (2003) showed that there is neither qualitative, nor quantitative difference in IRFs of Output and Inflation if one simulates a fully fledge model with either a time-contingent rule for firms' information updating (the SI Phillips Curve), or a time-contingent rule for being active (the Hybrid Phillips Curve). Luo and Young (2005) showed that more sophisticate models of Rational Inattention seem not to provide any relevant improvement to the dynamics of a general equilibrium model. Whereas, I challenged the SI Phillips Curve to explain 2 different features of the dynamics of Inflation – the positive correlation between the level and the persistence of Inflation, and the time varying volatility of Inflation. And I found comforting results.

The rest of the paper is organized as follows: in Section 2, I review the Sticky Information literature and MR model. Section 3 provides the econometric strategy and the results of GMM estimation. Section 4 is devoted to the analysis of those facts about the behavior of Inflation that can't be explained by New Keynesian Phillips Curve. Some conclusions are given in Section 5.

2 Beyond New Keynesian Phillips Curve: The Sticky Information hypothesis

In the last decade we saw the attempt of a good number of distinguished economists to impose the New Keynesian (henceforth NK) Synthesis as standard benchmark model for monetary economic practitioners; to say, a standard toolbox to perform monetary policy exercises or macroeconomic forecasts.

NK Synthesis reproduces quite well the dynamics of the economy under many aspects, even though strong critics arose because it misses to explain the stickiness observed in time paths of macroeconomic variables. Because of this weakness, its opponents derived a set of facts that the NK Phillips Curve fails to match up: why Inflation responds gradually to monetary policy shocks, Mankiw (2001); why output losses typically accompany a reduction in Inflation, Able and Bernanke (1998). Plus, they stressed some counterfactual implications of NK Phillips Curve: announced disinflation causes a boom, Ball (1994).

The challenge to square these facts gave the born to a number of alternative models, each based on a different assumption about the behavior of firms at micro level, whose theoretical support comes out from the optimal solution of profit maximization problem under different constraints on the set of feasible strategies for the firm. Those constraints may apply either on firms' *actions*, e.g. in some periods the firms are forced to be inactive and/or to follow suboptimal behaviors (like indexation), or on firms' *attention*, e.g. the firms face some constraints to absorb the relevant Information. It worth notice that both the hypotheses themselves, *Inaction with Indexation* and *Inattention*, reproduce quite well the observed persistence in aggregate dynamics of variables, as showed by Trabandt (2003), and Luo and Young (2005).

Regarding the Inaction hypothesis, worth a mention the literature that repropose NK Synthesis by allowing for ad-hoc Indexation-to-past versions of NK Phillips Curve that trigger structural stickiness into the Keynesian framework.[2] For what concern the Inattention hypothesis, the literature is based on the optimization problem of an agent with limited Information flows. Information constrains may be exogenous – as in Mankiw and Reis (2002) –, or related to a cost-benefit analysis – Reis (2004) achieved a microfundation of Sticky Information Phillips Curve based on costly information –, or because of physical constrains – Sims (1998, 2003) used Information Theory to model the Inattentiveness: i.e. a physical constraint to agents' Information channels; in this spirit, among the others, Moscarini (2004) solves the firms' profit maximization problem, finding an equation for the aggregate level of prices; Luo and Young (2005) simulate a fully fledge DSGE model with Rational Inattention à la Sims.

In particular, MR model takes the move from a reasonable and intuitive argument: what would be the outcome of a producers' model if one solves the expectations using a rule which is able to reproduce the main stylized facts about agents' expectations, as they are reported in surveys of consumers' expectations? Hereafter I first present the surveys of expectations which MR linger over to support their rule, and then I explain the setup of model.

2.1 A stylized fact in Expectations: the Disagreement

Using the microdata coming from the Surveys of Agents' Expectations,[3] Mankiw Reis and Wolfers (2003) underlined that the disagreement within a group of agents is one of the crucial characteristics of the expectations, both among Consumers and Professionals. In particular, they find that the interquantile range of Inflation expectations for 2003 among professional economists goes from 1.5 to 2.5 percent.[4] Among the general public, the interquantile range of expected inflation goes form 0 to 5 percent,[5] see Figure 1. Very similar findings are obtained in Carroll (2001), and pioneered in Roberts (1998), (2001).

[2] The interested reader may refer to Galì and Gertler (1999).

[3] The two Surveys analyzed was the Michigan Survey of Consumers' Expectations, and the Survey of Professional Forecasters.

[4] Predictions for 12-months-ahead Inflation in US (source Survey of US Professional Forecasters).

[5] Predictions for 12-months-ahead Inflation (source Michigan Survey).

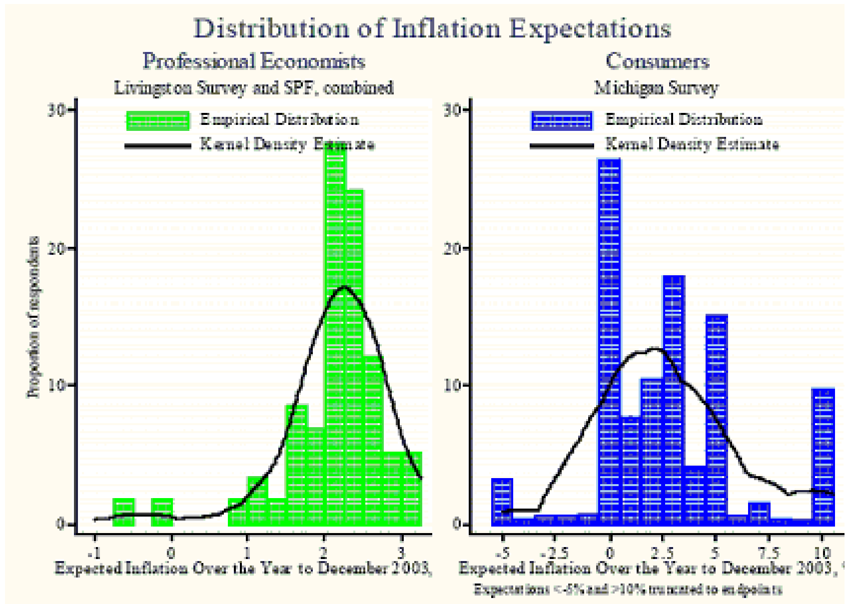

Figure 1. Cross-section distribution of Expectations (Source Mankiw et al., 2003)

Also, they provide evidences that the degree of rationality of expectation is neither fully rational, nor naive as in the adaptive expectations. In particular, the monthly time series of one-year-ahead Inflation forecasts shows the following properties:

1. No bias; in other words, the mean error on median prediction with respect to actual value for the period is around zero.
2. The information contained in the forecast is fully exploited; that is, the residual of forecast has no correlation with the forecast itself.
3. The forecasting errors are persistent, in the sense that realized last year prediction error has a positive correlation with today forecast error.
4. Not all the information available at time t is fully exploited for next periods forecasts made in t and after.[6]

These general findings pave a tiny way to depart from Rational Expectations (henceforth RE). Data match up with a rational processing of the information used (see above 1. and 2.). Yet, not ALL the information available is used (4.), and the positive correlation between errors (3.) suggests that there is information in last year forecasts that has not been exploited in generating this year's forecasts, thus violating the full rationality hypothesis.

[6] When I say *available at time t*, I mean that the information was in principle accessible in that period, not that all the variables dated t must belong to the Information at t. So, for instance, we may think that only variables dated $t-1$ and before are available at time t.

It's worth notice that the autocorrelation in forecast errors is not itself a test of rationality unless the errors are not overlapping, as argued by Batchelor (1982), because if the agents couldn't know the realization of the errors at the time of the forecast, then clearly these errors would add up in the forecast errors even under RE. This implies that, if agents are asked to forecast Inflation 12 months ahead, we should check for correlation between ε_t and ε_{t-12} and before (using monthly data).

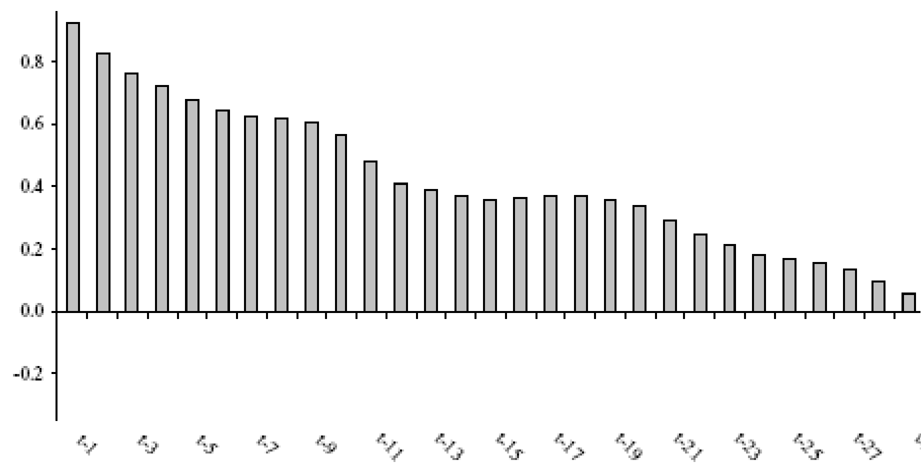

Figure 2. Expectation Errors: Autocorrelations (Source Kenny, 2002)

From Figure (**??**) is apparent that correlation is positive also beyond $t-12$, although it decays gradually over time. Such weakness of rational behavior may be explained allowing for a fraction of the agents to have adaptive expectations. Indeed, given that the inflation itself shows persistence, if the agent uses its past period value to perform the prediction, he will make a systematic error, which itself is highly correlated with the one he made in the period before. And the correlation will be higher the higher is the degree of persistence in Inflation.

2.2 Disagreement explained by Sticky Information

Having in mind these stylized facts, MR set a time contingent rule for agents' expectation, where every agent *processes the information rationally, but the rate at which he absorbs the relevant information is slower* than the one entailed by RE hypothesis. Consequently, some firms will form expectations conditional on a outdated Information Set, even when new information would be available.

In particular, MR assumed that only a constant fraction of the population gets new Information in the period, thus optimizing its decisions according with RE hypothesis, while the other firms choose their optimal actions using expectations conditional on an outdated Information. This means that they decided the best action when they last updated the information, and then they stick to their j period ago forecast about the best action to be taken today.

Mankiw at al. (2003) show that such rule does a pretty good job in reproducing the facts listed above, and it explains why almost all the surveys indicate that expectations across agents have some elements of rationality, but they are not fully rational.

There are various reasons for the agents to be *inattentive* to new information. MR loosely provide microfoundation. Yet, Reis (2004) support the inattentiveness hypothesis arguing that new information is costly, so that the agent will get new information only if the expected benefit is higher than the cost. Sims (1998, 2003) argued that what really binds in absorbing new information is the time a person devotes to thinking to macroeconomic conditions. He models the inattention hypothesis as if agents were interacting with the real world through a limited-capacity information channel, so that they can absorb a finite amount of informations in any period to reduce the uncertainty about relevant variables in decision processes.

Moscarini (2004) takes the insight of Sims addressing optimal time-dependent prices adjustment rules for firms. He generates optimal inertia from the frictions in the acquisition and/or processing of relevant information.

Branch (2004) explains the individual inattentiveness as a function of the increase in forecasts accuracy once new information is processed; that is, the more the updated set of information improves Thail index of forecasts (with respect to the outdated set), the more people will be attentive.

On the contrary, Carroll (2001) argues that the stickiness of information is implied in the way people gather macroeconomic news, i.e. the newspaper. Thus, the slow diffusion of news owns to the periodic emphasis that editors give to macroeconomic conditions. Using a model derived from theoretical epidemiology, Carroll tracks the spread of a piece of information as if he were tracking the spread of a disease through population, and the probability of the contagion is the one he assigns to agent's information updating.

2.3 The SI Phillips Curve

The baseline derivation of pricing decision rule in MR model is the standard monopolistic competition market populated by a continuum of *ex-ante identical*[7] firms in the (0,1) interval, with perfect competition on labor market.

In this setup, when the economy goes into a boom, each firm experiences an increased demand for its product. Because marginal cost rises with higher levels of output,[8] an higher demand means that each firm is likely to raise its

[7] When I say *identical* I mean that the firms have identical market power (demand elasticities are the same for each good), and identical ability to process the information.

Nonetheless they will not be identical ex-post, because some of them will update their information sets, thus setting the optimal prices, while the others will set prices conditional on outdated information.

[8] Under the assumptions of no variable capital, standard loglinearized relationship between marginal cost and output states that $mc_t = (\sigma + \varphi)y_t$, where σ is the inverse of consumer's intertemporal elasticity of substitution, and φ is the Fisher elasticity of labor supply.

relative price. In this case Woodford (2003) shows that the desired equilibrium price for the (identical) producers is given by

$$p_t^* = p_t + \alpha y_t \tag{1}$$

where α is a structural parameter of the economy. All the variables are expressed in log deviation from steady state. y_t is intended as the output gap.

Given that, MR assumed that every period a fraction $\lambda \in (0, 1]$ of the firms updates its information about the current state of the economy, and computes optimal prices conditional on that information. The rest of the firms continues to set prices conditional on outdated information. Each firm has the same probability to update its pricing plans, regardless of how long it has been since its last update.

So each firm *i,* which updated its plans j periods ago, adjusts today price accordingly to:

$$x_t^{j,i} = E\left[p_t^{*,i} \mid \Omega_{t-j}\right]$$

where $x_t^{j,i}$ is price adjustment of firm i at period t. All the variables are expressed in logs, and Ω_{t-j} is the information set at period $t - j$.

Since all the firms are ex-ante identical, the optimal price is the same for all those firms that have information dated *t-j.* So,

$$x_t^j = E\left[p_t^* \mid \Omega_{t-j}\right] \tag{2}$$

where we suppressed the firms index i.

Given the inattentiveness assumption, and the above structural relationship, MR derived a behavioral equation for the aggregate price level in the economy where the "short-run Phillips curve is apparent... output is positively associated with surprise movements in the price level".[9] In particular, let (1) to be the producer's desired price, and (2) to be the price adjustment in period t when the agent has information dated $t - j$. Then, MR showed that Inflation at time t will depends on contemporaneous Output, and on the expectations of Inflation and Output growth.

The dynamics of Inflation in this exercise is given by the following equation:

$$\pi_t = \frac{\alpha\lambda}{1-\lambda} y_t + \lambda \sum_{j=0}^{\infty} (1-\lambda)^j E\left[\pi_t + \alpha \Delta y_t \mid \Omega_{t-1-j}\right] \tag{3}$$

where $\Delta y_t = y_t - y_{t-1}$ is the growth rate of output gap, and λ is the probability that the agent updates his information in each period.[10] Thus, today's Inflation is affected by the expectations formed in all the past periods, with a weight that fades out at the rate $(1 - \lambda)$.

As claimed by its authors, equation (3) triggers persistence in Inflation dynamics because Inflation at period t depends on all the past periods expectations

[9] Mankiw and Reis (2002) pg. 1300
[10] For the proof and the details see Mankiw and Reis (2002).

that the producers have had on today's Inflation and Output growth. The explanation of the mechanism behind is the following: suppose that in t a shock occurs that increases the Output gap. Accordingly to (3), the Inflation starts raising from period t because of the trade off term $\frac{\alpha\lambda}{1-\lambda}y_t$. In the following period $t+1$, a fraction λ of the agents gets aware of the shock, so that Inflation raises again because now $E\left[\Delta y_{t+1} \mid \Omega_{(t+1)-1-j}\right]$ is greater than 0 for those agents with $j<1$. The same will happen in $t+2$, where now a fraction $\lambda(1-\lambda)$ gets aware of the shock, and so on for all the following periods, with an effect of expectations on Inflation that fades out at the rate $(1-\lambda)^j$.

2.4 Estimates of SI Phillips Curve

As I said, the main purpose of SI Phillip Curve was to solve the counterfactual implications of NK Phillips Curve. In their paper MR (2002) performed this task by simulating the model. They, indeed, showed that SI Phillips Curve responds gradually to Monetary Policy shocks, and its time path is highly persistent. Reis (2004) showed also that SI Phillips Curve was able to fit actual data of Inflation "remarkably well".

Both the papers achieved the results using a very specific calibration of the Information stickiness parameter, $\lambda = 0.25$. This value implies that the agents on average update their information sets once a year, which is an evidence they draw from the surveys on Inflation Expectations.[11] Unfortunately, micro evidences not always fit macro models. For example, in the case of Fisher labor supply elasticity, the value one needs for Real Business Cycle model to fit labor volatility is something close to 1, whereas the Fisher elasticity that comes out from surveys is around 1/6. Moreover, a working paper from UPF[12] shows that taking survey expectations to fit the Phillips Curve is misleading. Therefore, an empirical analysis is necessary in order to evaluate the model.

There is only another estimation of SI Phillips Curve I'm aware of: Khan and Zhu (2002). They estimate equation (3) for U.S., Canada, and U.K, finding values of lambda around 0.3, so almost in line with the calibration used by MR. To estimate (3) they proceeded as follows: first, they truncated at $t - j_{\max}$ the infinite sum of expectations; then they worked out the expectations by substituting them with the predictions taken from a VAR set ad-hoc to minimize the RMSE, since this should be the best linear counterpart of RE.

The estimation strategy of Khan and Zhu is criticizable for two reasons. The first one is technical: they used estimates as regressors without adjusting the standard errors to take it into account. And this is not a minor correction. Khan and Zhu used various truncation thresholds to check for robustness of their estimates, up to $j_{\max} = 11$. This means they included variables up to the 11-periods-ahead VAR forecasts among the regressors. Those familiar with VARs literature know that the variance of forecasts grows higher the further ahead you take the forecasts. Now, in this paper I used 1-period-ahead VAR

[11] See Mankiw et al. (2003)

[12] Nunes (2006).

predictions among the regressors. The adjusted standard error was 29% higher than before. Therefore, I reckon that the confidence interval for the estimates in Khan and Zhu paper should be much bigger than what they claim, and their estimation turns out to be very imprecise.

The second criticism regards the econometric strategy: Khan and Zhu estimate a difference equation of Inflation. Basically, they exploit the persistence in data to achieve an estimation of the Sticky Information parameter λ. Now, clearly the closed form they achieved comes from MR model. Nonetheless, it can be shown that such closed form holds true even when the DGP is another model, i.e. the sticky price model.

Moreover, if the agents in the DGP can either be *Inattentive*, or use *Indexation rules*, which both are sources of persistence in Inflation, one may find a strong upward bias in the coefficient of Information stickiness by estimating a model that address all the persistence to the inattentiveness source. This strategy seriously affects the estimation of SI coefficient if the data came from a model with heterogeneous agents - inactive, inattentive, and adaptive - as our intuition suggests.

2.5 Why not to use the first moment of SI Phillips Curve

I argued above that we should not attempt an estimation of first moment of Inflation if our purpose is to test the SI Phillips Curve against the alternative model of Sticky Prices, because this test turns out to have very low power.

Such claim is based on the empirical observation of the two models. In particular, notice that the New Keynesian Phillips Curve is

$$\pi_t = \gamma_1 \pi_{t-1} + \gamma_2 y_t + \gamma_3 E_t [\pi_{t+1}] \tag{4}$$

where $\gamma_1, \gamma_2, \gamma_3$ are the reduced form coefficients.

Under some sensible assumptions, it can be shown that SI Phillips Curve can be written as[13]

$$\pi_t = \beta_1 \pi_{t-1} + \beta_2 y_t + \beta_3 \Delta y_t + \beta_4 \Delta y_{t-1} \tag{5}$$

where $\beta_1, \beta_2, \beta_3, \beta_4$ are positive coefficients functions of the structural parameters α, λ and ρ.

Now, in Galì and Gertler (1999) it is pointed out that Output gap leads Inflation in actual US data; that is, the higher (positive) correlation is between contemporaneous Output and one-step-ahead Inflation. This means that when we observe an increase in Output – or equivalently a positive growth of Output – at period t, we expect higher Inflation in the following period $t+1$. Hence, if we are to estimate equation (5) with US data, we will find significant coefficients for β_3 and β_4 either when (5) is actually the DGP, or when (4) is the DGP. In this last case, indeed, β_3 will be significant because it captures the positive correlation given by the $cov[\Delta y_t, \pi_{t+1}]$, while β_4 captures the $cov[\Delta y_{t-1}, \pi_t]$.

[13] See Appendix A for the derivations. I use there almost the same assumptions used by Makiw and Reis (2002) to simulate the model.

So, we may find an overall good result for the estimation of (5) even when the model is false, Type II error, which is the point I complained about using the first moments of SI Phillips Curve.

3 The Estimation

To overcome the problem of low power against the alternative of sticky prices I derive and estimate the theoretical second moment of Inflation from SI Phillips Curve. This equation is valid only in SI model, and not in sticky prices one. Thus, it is able to disentangle between the two theories.

3.1 Econometric specification

Assume that the dynamics of Inflation and Output growth in the economy is given by the interaction of n variables, which are the elements of some covariance stationary stochastic vector process Z_t. Although this assumption ain't very restrictive and it poses very few structure on the Inflation and Output process, it allows us to write down analytically the theoretical second moments of Inflation from SI Phillips Curve.

Lemma 1 *Let* $\{Z_t\}_{t=0}^{\infty}$ *be a covariance stationary* $(n \times 1)$ *vector process s.t.* $\{\pi_t, \Delta y_t\} \subset Z_t$*. Then equation (3) implies:*

$$\frac{\alpha\lambda}{1-\lambda} y_t + \alpha \Delta y_t = \sum_{i=0}^{\infty} (1-\lambda)^i \, \delta A_i \varepsilon_{t-i} \tag{6}$$

where the $(n \times n)$ *matrices* A_i *are the dynamic multipliers of the* Z_t *process, and* ε_t *is the* $(n \times 1)$ *vector process of the exogenous shocks.* δ *is a* $(1 \times n)$ *row vector that picks up* $(\pi_t + \alpha \Delta y_t)$ *within* Z_t *vector.*

Proof. See Appendix B. ∎

Once I obtained equation (6), it's straightforward to achieve an econometric specification for the model. In particular, I choose the following orthogonality condition: multiply equation (6) by $(\delta\varepsilon_t)'$ and take the expectations at time t, then:

$$E\left[\left(\frac{\alpha\lambda}{1-\lambda} y_t + \alpha \Delta y_t\right)\left(\alpha\varepsilon_t^{\Delta y} + \varepsilon_t^{\pi}\right) - \delta\Sigma\delta'\right] = 0 \tag{7}$$

where $\Sigma \equiv E\left[\varepsilon_t \varepsilon_t'\right]$ is the Variance-Covariance matrix of the shocks.

Orthogonality condition (7) is handy for the estimations, since it works out the infinite summation in (6). Plus, it will be useful to investigate the effect of SI hypothesis on the variance of Inflation, which is the object of section 4.2.

Next step is to estimate the sample analog of (7) using the GMM. Such sample analog, however, would have the true shocks ε_t and the population variance Σ as regressors, which are unknown to us. Nonetheless, it can be shown that if one uses $\widehat{\varepsilon}_t^{\Delta y}, \widehat{\varepsilon}_t^{\pi}, \widehat{\Sigma}_T$ as regressors, the sample analog as well converges almost surely to the population moment (7), given that these regressors are consistent estimators of $\varepsilon_t^{\Delta y}, \varepsilon_t^{\pi}, \Sigma$.

Hence, I estimate (7) where the candidate regressors – the residuals $\widehat{\varepsilon}_t^{\Delta y}$, $\widehat{\varepsilon}_t^{\pi}$ and the Variance-Covariance matrix $\widehat{\Sigma}_T$ – come from the estimates of a $VAR(p)$, where p is the minimum order of lags in order not to have autocorrelation in the VAR residuals.[14]

For robustness analysis, I estimated various specifications of the VAR, with different numbers of dependent variables, from a minimum order specification $\{\Delta y_t, \pi_t, i_t\}$[15] to a maximum order specification, which is intended to minimize the $RMSE$.[16] The complete set of variables in this last case includes: either Output gap, or Output Gap Growth; Inflation measured as CPI, or as the Implicit GDP deflator; Short-term Interest Rate (the Fed Fund Rate); Long-term Interest Rate (10 years Gov't bond rate), or the Term Spread (long run minus short run interest rate); first difference or HP filter of real Stock Price Index (S&P500, deflated by CPI); first difference or HP filter of Price Index of Commodities; first difference or HP filter of Real Money (real M2 minus small time deposits); Unemployment Rate; Total Capacity Utilization Rate.[17] All the variables are in logs except for the unemployment, total capacity utilization, and interest rates rates.

The estimation procedure is the following: first, I estimated the VARs and I saved the results; second, I used $\{\varepsilon_t(\beta), \Sigma_T(\beta)\} \mid_{\beta=\beta_T^{VAR}}$as regressors in orthogonality condition (7) to obtain λ_T^{Gmm}.

3.2 Results

I estimate (7) with GMM, which in this case is the Non-linear IV estimator, but with a smaller variance. I use a set of 19 instruments,[18] all dated $t-1$ and before, which in principle are included in the information set available at time t, so being uncorrelated with the GMM residuals. All the variables, both for GMM and VAR estimation, come from FRED II database of US economy.[19] The variables are in logs, they have been detrended or differenced when necessary.

[14] Autocorrelation of residuals is tested with a standard LM test.

[15] I follow here Cogley and Sargent (2005).

[16] In details, the process Z_t is

$$\underbrace{Z_t}_{n\times 1} = \begin{bmatrix} \Delta y_t & \pi_t & X_t' \end{bmatrix}'$$

where X_t can be either $X_t = i_t$ or a $(n-2\times 1)$ vector that includes the set of variables which minimize the $RMSE$ in the equations of Output growth and Inflation.

[17] See Stock and Watson (2003) for recent and very exaustive assess about forecasting Inflation and Output. See also Sims and Zha "Macroeconomic Switching" (1996) for some interesting issues on this argument.

[18] In details, I used a constant plus 4 lags of inflation and output gap, two lags of unemployment rate, interest rate, marginal cost, money growth, and term spread.

[19] Available at the Federal Reserve Bank of St. Louis.

All the variables used in the VAR are stationary, and the VAR residuals are checked to be not serially correlated in all the specifications. The sample used to estimate the VAR is 1957q1 to 2005q4, whereas the GMM uses a smaller sample – 1958q4 to 2005q4, (189 obs.) – since I lost 7 lags to estimate the VAR.

To control for small sample bias that affect nonlinear GMM estimations, I estimate two alternative specifications of (7). Specification one is (7) times $(1-\lambda)$; specification two is (7) times $\frac{1-\lambda}{\alpha\lambda}$. They are referred to as (1) and (2).

The estimation results are summarized in the following Table 1.

Table 1.

Restricted $\alpha = .2$	Spec	λ^{Gmm}	Adjust Std.Err	Null 1: MR cal.	T-stat (p-val)	Null 2: RE	T-stat (p-val)	J-stat (p-val)
VAR $\{\Delta y_t, \pi_t, i_t\}$	(1)	.751	.0664	0.25	7.55 (.000)	1.00	-3.74 (.000)	22.20 (.224)
	(2)	.864	.0679	0.25	9.05 (.000)	1.00	-1.99 (.046)	15.29 (.641)
VAR *min RMSE*	(1)	.860	.0829	0.25	7. 36 (.000)	1.00	-1.68 (.092)	21.99 (.232)
	(2)	.911	.0840	0.25	7.88 (.000)	1.00	-1.04 (.294)	14.65 (.685)

US data, sample 1958q4 – 2005q4. HP filter for output gap. NW Standard Errors, adjusted for stochastic regressors. J statistic from Hansen test of overidentifying restrictions (19 moments). 2-step GMM with optimal weighting matrix.

Overall, the estimates are reasonable and quite precise, and the model fits well the data accordingly to the J statistics (Hansen Test). We can never reject the null hypothesis of overidentifying restrictions.[20]

I estimate only λ, while I didn't attempt to estimate α. This last is a structural parameter of the economy, which depends on the intertemporal elasticity of substitution of the consumers, on Fisher elasticity of labor, and on the elasticity of the demand of the single goods. Since I use a Limited Information estimation – in the sense that I estimate not a full model of the economy, but just an equation that holds on the supply side – it seemed to me pointless and dangerous to force the estimation of a parameter that depends on equations that don't belong to the specification used.

[20] A yellow flag is laid here. When using Instrumental Variables estimators, is crucial to check for the presence of weak intruments, otherwise one cannot fully rely on hypothesis testing.

Unfortunately, how to check for weak instruments is still unclear in the literature when the estimation has possibly non-spherical residuals, and/or non-linear orthogonality conditions. An hypothesis testing robust to weak instruments might be the object of an extension of present work.

About the estimates of the degree of Information stickiness, in all the specifications tested λ_T^{Gmm} is in the range predicted by the theory, within the right interval $(0, 1]$; more precisely, it is always within the range $[0.75, 0.95]$. This values imply that the average frequency of information updating of the firms is around a quarter and a month. As a matter of fact, this value is far away from the standard value used in calibrations, e.g. MR (2002) or Reis (2004). This is not surprisingly: MR choose $\lambda = 0.25$ in order to make SI able to trigger the degree of persistence observed in actual data, *as if* SI were only source of persistence. But I explained above how misleading is such exercise if the DGP is a model with heterogeneous agents where SI is only one of the sources of persistence. Once we isolate the effect of SI alone, my estimates tell us a quite different story from the one told us about the authors of the SI model.

To assess empirically MR calibration, however, I test the Null Hypothesis of $\lambda = 0.25$. In all the cases, the null is rejected. My conclusion is that such value is not conformable with actual data for US economy.

In some of the specifications tested, however, λ_T^{Gmm} is quite close to the RE value 1. Therefore, my second concern has been to test whether the Rational Expectations hypothesis was accepted by the data. In this case the results are not overwhelming. The Null is rejected at 5% level in all the minimum order specifications, but it is accepted in the ones that minimize $RMSE$.

My feeling about these results is that there is some room for SI to explain actual data, even though it is far less important that what their authors initially claimed.

3.3 Robustness analysis

3.3.1 Empirical robustness

I checked the robustness of the results along five dimensions:

1. Different specifications of VAR model, as explained in previous section;
2. Sensitivity to the calibration of α;
3. Various specifications of the Z_t process: AR(2), AR(p);
4. A different filter for the gap variables: the Band Pass (BP) instead of the Hodrick Prescott (HP);
5. A different orthogonality condition, (7'), derived from (6) (see Appendix C.)

I've already presented the results for the different specifications of the VAR. The following Table 2. shows the estimations in the other 4 cases.

Table 2.

Robustness Analysis	λ^{Gmm}	Adjust Std.Err	Null Mr cal: $\lambda^{Gmm} = .25$	Null RE: $\lambda^{Gmm} = 1$	J-stat (p-val)
$\alpha = .1$	.937	.0503	13.67 (0.00)	-1.23 (0.21)	13.44 (.764)
Z_t~AR(2)	.635	.1821	2.11 (0.03)	-2.00 (0.04)	6.38 (.172)
BP filter	.912	.0700	9.46 (0.00)	-1.24 (0.21)	18.30 (.435)
Orthogonality Condition (7')	.799	.0102	53.68 (0.00)	-19.65 (0.00)	18.75 (.281)

US data, sample 1958q4 – 2005q4. HP filter for output gap except than in the 3rd case. NW Standard Errors, adjusted for stochastic regressors. J statistic from Hansen test of overidentifying restrictions (19 moments). 2-step GMM with optimal weighting matrix.

The empirical robustness tells us three thing:

- It confirms the results found in Table 1. Also, the tests of the two Nulls give the same qualitative answers. So, the estimation seems quite robust, and the conclusions stated above do not change.
- I found that the smaller are the errors $\widehat{\varepsilon}_t$, the bigger is λ_T^{Gmm}, which is a finding that makes sense.
- It uncovers a stable *inverse* relationship among the value of α and the one of λ, which is true in all the specifications.

3.3.2 Theoretical robustness

One reason of discontent in using orthogonality condition (7) is that it doesn't exploit the information about λ set in the RHS of (6). It is an efficiency cost I must bear to achieve an econometric closed form out of (6).

As robustness exercise, however, I simulate here the implications of λ on the theoretical second moments when exploiting both LHS and RHS of equation (6), in order to see whether the estimation based on (7) is confirmed.

Multiplying equation (6) by itself, and taking the expectations we obtain the second moment of (6):

$$E\left[\frac{\alpha\lambda}{1-\lambda}y_t + \alpha\Delta y_t\right]^2 = \sum_{i=0}^{\infty}(1-\lambda)^{2i}\,\delta A_i \Sigma A_i' \delta' \tag{8}$$

Looking at (8) the differences between the two moments (7) and (8) are apparent. In (7) all the terms of the summation cancel out except for the first one. Consequently, λ disappears from the RHS of the equation, and when I estimate the equation, it can only exploit the information about λ contained in the LHS.

In (8), instead, λ remains as a coefficient of the summation, which determines the rate of convergence.

In order to exploit this double source of information about λ, in (8) I computed both the sample mean of LHS using the actual data, and the RHS using the variance of the shocks $\widehat{\Sigma}$ and the dynamic multipliers $\widehat{A}_i$ that come from the same VAR used above. Notice that a solution of (8) in $\lambda \in (0,1)$ must exist for some values of alpha; that's because, on one end of the interval, for $\lambda \to 0$, we have in LHS: $E\left[\frac{\alpha\lambda}{1-\lambda} y_t + \alpha \Delta y_t\right]^2 \to \alpha^2 E(\Delta y_t^2)$, whereas the RHS becomes a linear combination of the variance of the Z_t process, i.e. $\delta VCV(Z_t)\delta' > 0$. Thus, there always exists a $\bar{\alpha} > 0$ s.t. for any $0 \leq \alpha \leq \bar{\alpha}$ holds $LHS < RHS$.

On the other end of the interval, for $\lambda \to 1$, the LHS goes to plus infinity because of the first term, whereas the infinite summation in RHS collapses to its first term, $\delta\Sigma\delta'$; thus, when λ increases after some point it will hold $LHS > RHS$ for any value of alpha chosen before. Therefore, for any $\bar{\alpha}$ there exist a $\lambda \in (0,1)$ s.t. $LHS = RHS$.

Figure (3.) illustrates the result for this exercise.

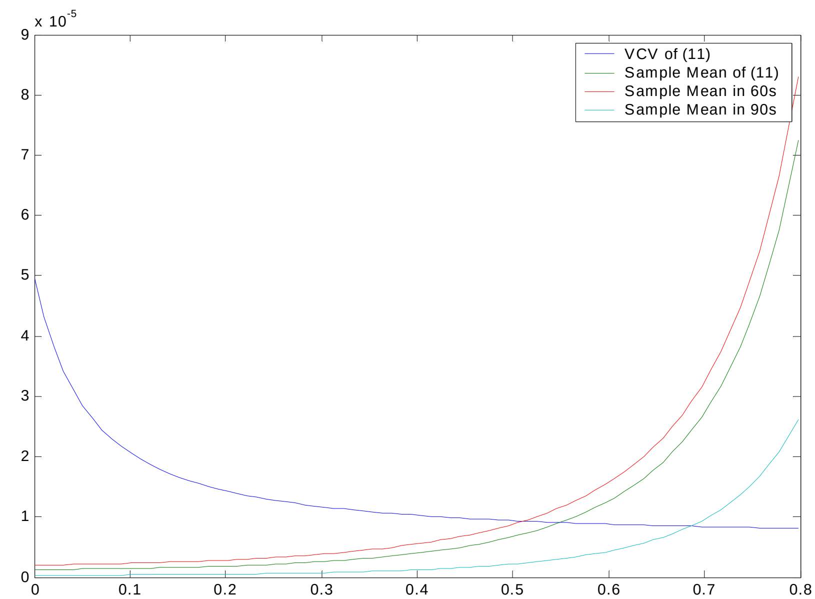

Figure 3.

As we expected there is a point of intersection, but the estimate of λ here differs substantially from the one in the previous estimations. With the calibration of α used above, (8) holds for $\lambda \approx .5$. Therefore, either present estimation of λ is downward biased, or the one we get from (7) is upward biased. Further investigations are needed here.

4 Sticky Information and Inflation Dynamics

Are the conclusions attained above valid overall the sample? Has SI always been almost irrelevant for Inflation persistence? Plus, are there some other aspects of the Inflation dynamics, beyond the persistence, that can be explained by SI? In other words, can we make some qualifications to the above conclusions? My answer is yes, and I'll present two exercises in this spirit.

First, I'll check whether the coefficient of SI has had structural breaks during the considered sample. Actually, I'll show that in the 60's the degree of information stickiness was much higher, so being a more important source of Inflation persistence.

Second, I'll investigate the implications of SI on the volatility of Inflation. I find that the SI coefficient has an effect that either magnify or buffer the variance of Inflation shock, $var\left(\varepsilon_t^{\pi}\right)$, with respect to the variance of the exogenous innovation, depending on the sign of the covariance between the shocks on Output gap growth and Inflation. Hence, SI can explain why we might observe different variances of Inflation in different periods, even when the variance of the exogenous shocks is constant. In other words, SI gives a different explanation of the time varying volatility of Inflation, which is opposite to the one that rely on stochastic volatility of the shocks.

It is worth notice that this feature of SI model it is not present in the Hybrid Phillips Curve, where the backwardness parameter has a monotone (and positive) effect on the variance of Inflation. In this fashion, the SI model would improve our ability to reproduce (endogenously) the dynamics of Inflation beyond what a Phillips Curve with leads and lags of Inflation can do.

4.1 Structural Breaks Tests

This section focuses on the link between Inflation persistence and Information stickiness during the sample. The issue is interesting, because for the alternative model of sticky prices it is an empirical conundrum the fact that the disinflation of the 90's was accompanied by an increase in the coefficient of forward-looking Inflation Expectations; in other words, a fall in inflationary persistence.[21] The argument is that "It's difficult to see why a reduction in inflation and inflationary uncertainty would be accompanied by lower persistence in a model relying only on staggered contracts or menu costs to explain nominal inertia. Lower and more stable inflation would seem to be a force for lengthening contracts, implying greater persistence in inflation..."[22]

On the contrary, the inertia in SI model depends on the level of uncertainty about the economy - the more relevant are the unpredictable shocks to aggregate demand, the more persistent will be the Inflation. And it makes sense to think that in recent years the firm have had a better knowledge of the economic system than in the past. Therefore, my purpose here is to see whether the Information stickiness parameter diminished – i.e. λ increased – during the disinflationary

[21] The point was made in Erceg and Levin (2003) and Bayoumi and Sgerri (2004).
[22] Bayoumi and Sgerri (2004) pg. 6

period of the 90's. Since the econometric specification matches the theoretical second moment with the data on Inflation volatility, λ is estimated without any link to the degree of persistence in the data. Therefore, there is in principle no reason why we should observe a structural change in λ to match the change in persistence, which makes the test interesting and not trivial.

I perform Andrews (1993) supLM test of structural breaks.[23] It cuts the tails of the sample and computes the most likely point in time where a break might have occurred in the remaining middle subsample. I summarized the results in Table 3. I run the test for various specifications, and I report the range, the mean, and the median of the supLM statistics.

Table 3.

Structural Breaks Tests **H_0: No SB**		Range for the 16 specifications	Mean (Median)	Asy Critic Val. 1%	Asy Critic Val. 5%	Asy Critic Val. 10%
Andrews supLM (1 parm)	$\pi_0 = .2$	{5.31,…,11.92}	7.27** (7.06**)	11.69	8.45	6.80
	$\pi_0 = .1$	{6.66,…,32.08}	16.21 (9.94*)	12.69	9.31	7.63
Subsamples Comparison $\lambda_{60} = \lambda_{90}$	Wald	{.11,…,61.7}	9.42 (1.19***)	Wald and LM have std. distr. $\chi^2(1)$		
	LM	{1.52,…,1871}	786.1 (57.72)	6.63	3.84	2.70

π_0 indicates the percent of tail cut. The supLM test has a non standard distribution. The asymptotic critical values are given in Andrews (1993).

The evidence of structural breaks is not overwhelming. The null hypothesis of structural stability is accepted at 5% both in mean and median for $\pi_0 = 0.20$, but it is rejected it if we use a larger sample ($\pi_0 = 0.10$). I noticed that when the test rejects the null, it places the break around the end of the 60's. From a closer look at the GMM residuals the reason of such result is clear. During the 70's the increase in the volatility of Inflation biases the results, and the test captures the spurious effect of the oil shock in the 70's as a structural break in λ – see Figure (4.)

[23] I choose the supLM because is the most powerful test when the timing of the (possible) break is unknown.

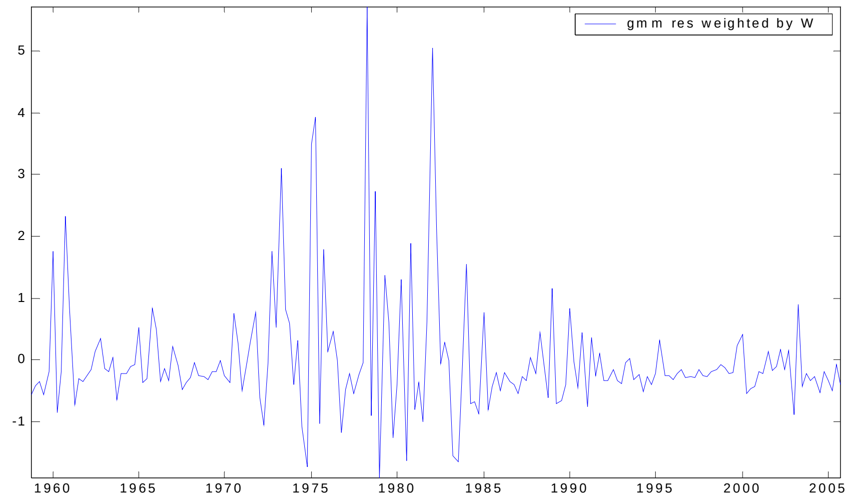

Figure 4.

Therefore, I perform a second set of tests to get rid of exogenous volatility bias in the 70's. Under the Null hypothesis that the same model holds throughout the sample, I compared the coefficients estimates in two subsamples of equal length: one that goes from 1959q1 to 1970q1, and the second from 1990q4 to 2005q4. The results are in Table (3.) above. The Null hypothesis that $\lambda_{60}^{Gmm} = \lambda_{90}^{Gmm}$ it is rejected in almost always.

These findings make sense. In past years, the Information was likely to be more sticky: less media, less accurate forecasts and previsions, less experienced authorities, less data gathering, etc. So, SI contributed to render stickier the responses of the firms to the events that happened at contemporaneous time. This transferred to a more persistent behavior of aggregate prices, i.e. of the Inflation. On the contrary, in recent years SI becomes a less important issue, and the contribution to persistence from this source becomes nonessential.

4.2 Volatility Analysis

Equation (7) has an intuitive interpretation as a restriction in the identification matrix of a structural VAR. In other words, the theoretical second moment of SI model implies restrictions on contemporaneous correlations between shocks. This feature of the model is more apparent if we write (7) as:

$$\sigma_{\pi}^2 = \frac{\alpha\lambda}{1-\lambda} cov(y_t, \varepsilon_t^{\pi}) - \alpha\sigma_{\pi,\Delta y} + \frac{\alpha^2\lambda}{1-\lambda} cov(y_t, \varepsilon_t^{\Delta y}) \tag{9}$$

The interesting thing here is that the restriction (9) is not constant in time, but it depends on the covariance between the shocks of Output and Inflation:

the more they covary negatively, the more they average out, and the more SI will reduce the variance of Inflation with respect to the variance of some exogenous innovation. To better see the point, take the derivative of σ_π^2 w.r.t. SI coefficient,

$$\frac{\partial \sigma_\pi^2}{\partial \lambda} = \frac{\alpha}{(1-\lambda)^2} cov(y_t, \varepsilon_t^\pi) + \frac{\alpha^2}{(1-\lambda)^2} cov(y_t, \varepsilon_t^{\Delta y})$$

It is clear that an increase in the degree of stickiness of Information, i.e. a reduction in λ, increases the variance only when the following statistic (10) is positive

$$\begin{aligned} &\frac{\partial \sigma_\pi^2}{\partial \lambda} \leq 0 \\ iff \quad &-\frac{cov(y_t, \varepsilon_t^\pi)}{cov(y_t, \varepsilon_t^{\Delta y})} \geq \alpha \end{aligned} \tag{10}$$

In the opposite case, where (10) holds with the less sign, SI buffers the exogenous variance.

In Figure (5.) I plot the annual means of the statistic (10) for the considered sample (blu line). The green line is $\alpha = 0.2$

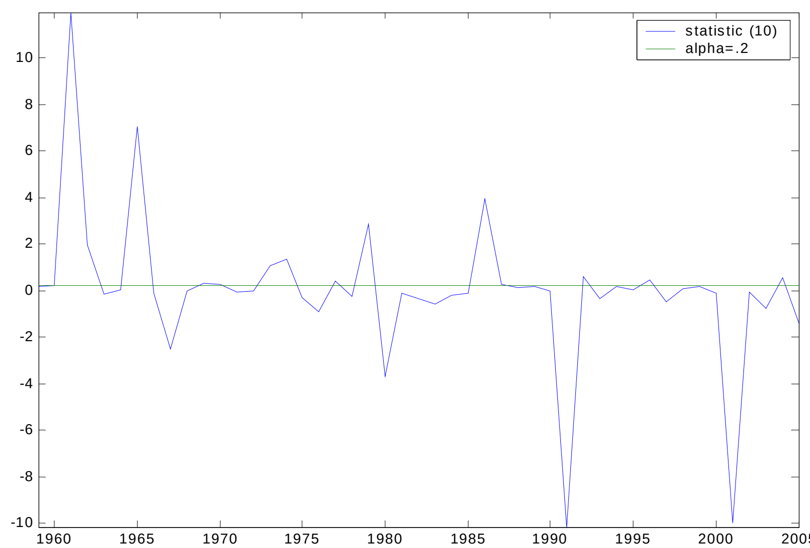

Figure 5.

In those years when the literature place the so-called "Great Moderation",[24]

[24] The "Great Moderation" is a substancial reduction in the volatility of the main macroeconomic aggregates that began in mid 80's. It has been first identified and named by Stock and Watson (2003).

the statistics (10) was negative, implying that SI operated in reducing the Inflation variance. (see Figure 5.)

I interpret this evidence as one of the causes of the reduction of the Inflation volatility observed in the 90's. Such reduction would be due not to stochastic volatility of the exogenous shocks, as claimed by Cogley and Sargent (2003) or Canova et al (2005), but to the particular mechanism that operates in a SI model.

5 Conclusions

The Information stickiness seems not to be a crucial issue to explain the inertia of Inflation. Moreover, in my estimations λ is more sensible to the accuracy of the forecasts,[25] than to the chained structure of the errors, which is the trademark of SI. Given my estimates – $\lambda \simeq 0.8$ – the contribution of SI to persistence of Inflation is small, and SI would be a negligible improvement to the standard NK framework, at least regarding the time path of Inflation. For this reason, thinking and setting up a model where the agents can be either Inactive or Inattentive, and then putting some effort to estimate it, overall these costs are too high compared with the benefits you get from such model in terms of predictability of the aggregate variables. This is the main insight we should get from present work, and it is a conclusion in line with Trabandt's or Luo and Young's papers.

In a different perspective, Sticky Information Phillips Curve is an interesting model to understand the dynamics of Inflation regarding the time varying volatility, which has recently become a debated issue in the literature. In particular, a more sophisticated model, with some version of SI or RI hypothesis, could be helpful to identify a series of the exogenous shocks, which are dependent in time and net of Inattentiveness component.

[25] Recall what I found in section 3.3.1: the bigger the errors, the smaller λ.

References

[1] Abel, Andrew and Bernanke, Ben (1998): "Macroeconomics", 3rd ed. MA: Addison-Wesley

[2] Andrews, Donald (1993): "Test for Parameter Instability and Structural Change With Unknown Change Point", Econometrica vol. 61 n^o 4, pp. 821-856

[3] Ball, Laurence (1994): "Credible Disinflation with Staggered Price Setting", American Economic Review 84, pp. 282-289

[4] Bayoumi, Tamim and Sgherri, Silvia (2004): "Monetary Magic? How the Fed Improved the Flexibility of the U.S. Economy", IMF wp 04-24

[5] Branch, A. William (2004): "Sticky Information and Model Uncertainty in Survey Data on Inflation Expectations", Unpublished Manuscript

[6] Canova, Fabio and Gambetti, Luca and Pappa, Evi (2005): "The Structural Dynamics of Output and Inflation: What Explains the Changes?", forthcoming Journal of Money Credit and Banking

[7] Carroll, Christopher (2001): "The Epidemiology of Macroeconomic Expectations", NBER wp 8695

[8] Cogley, Timothy and Sargent, Thomas (2005): "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US", Review of Economic Dynamics n^o 8

[9] Erceg, Christopher and Levin, Andrew (2003) "Imperfect Credibility and Inflation Persistence", Journal of Monetary Economics vol. 50, pp. 915-44

[10] Fischer, Stanley (1977): "Long-term Contracts, Rational Expectations, and the Optimal Money Supply Rule", Journal of Political Economy vol. 85 pp. 191-205

[11] Galí, Jordi and Gertler, Mark (1999): "Inflation dynamics: A structural econometric analysis", Journal of Monetary Economics 44, pg. 195-222

[12] Khan Hashmat and Zhu Zhenhua (2002): "Estimates of the Sticky-Information Phillips Curve for the United States, Canada and the United Kingdom", Banque du Canada wp 2002-19

[13] Lucas, Robert E. Jr. (1973): "Some International Evidence on Output-Ination Tradeoffs", American Economic Review 63, pp. 326–334

[14] Luo, Yulei and Young, Eric (2005): "Rational Inattention and Aggregate Fluctuations", manuscript Princeton University

[15] Mankiw, N. Gregory (2001): "The Inexorable and Mysterious Tradeoff Between Inflation and Unemployment", Economic Journal vol. 111

[16] Mankiw, Gregory and Reis, Ricardo (2002): "Sticky Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve", Quarterly Journal of Economics vol.117 (4), pg. 1295-1328

[17] Mankiw, Gregory and Reis, Ricardo and Wolfers, Justin (2003): "Disagreement about Inflation Expectations", NBER wp 9796.

[18] Moscarini, Giuseppe (2004): "Limited Information Capacity as a Source of Inertia", Journal of Economic Dynamics and Control, pp. 2003-2035

[19] Reis, Ricardo (2004): "Inattentive Producers", Manuscript Harvard University

[20] Sims, Christopher A. (1998): "Stickiness", Carnegie-Rochester Conference Series on Public Policy 49, pp. 317-356

[21] ——————— (2003): "Implications of rational inattention", Journal of Monetary Economics n^o 50, pp. 665-690

[22] Stock, James and Watson, Mark (2003): "Forecasting Output and Inflation: The Role of Asset Prices", Journal of Economic Literature vol. XLI, pp. 788-829

[23] ————————————- (2005): "Has the Business Cycle Changed? Evidence and Explanations", forthcoming FRB Kansas City symposium, Jackson Hole, Wyoming, August 2003

[24] Taylor, B. John (1980): "Aggregate dynamics and staggered contracts", Journal of Political Economy n^{o}88, pp. 1-23

[25] Trabandt, Mathias (2003): "Sticky Information vs. Sticky Prices: A Horse Race in a DSGE Framework", Manuscript Humboldt University Berlin

[26] Woodford, Michael (2001): "Imperfect Common Knowledge and the Effects of Monetary Policy", NBER wp 8673

[27] ——————— (2003): "Interest and Prices", Princeton University Press

A Derivation of Equation (5)

Define

$$S_t = p_t + \alpha y_t \tag{11}$$

and assume that the growth of S_t follows an AR(1) process

$$\Delta S_t = \rho \Delta S_{t-1} + u_t \tag{12}$$

where $u_t \sim N\left(0, \sigma_u^2\right)$ *i.i.d.*

Now, using the definition (12) and equation (12) into (3) I obtain:

$$\pi_t - \frac{\alpha\lambda}{1-\lambda} y_t = \lambda \sum_{j=0}^{\infty} (1-\lambda)^j E_{t-1-j}\left[\Delta S_t\right] \tag{13}$$

Given (12), we have that $E_{t-1}\left[\Delta S_t\right] = \rho \Delta S_{t-1}$, $E_{t-2}\left[\Delta S_t\right] = \rho^2 \Delta S_{t-2}$, and so on. Thus, I can work out the expectations in (13) to obtain:

$$\pi_t - \frac{\alpha\lambda}{1-\lambda} y_t = \lambda \sum_{j=0}^{\infty} (1-\lambda)^j \rho^{j+1} \Delta S_{t-1-j} \tag{14}$$

Taking out of the summation the first term I have

$$\pi_t - \frac{\alpha\lambda}{1-\lambda} y_t = \lambda \rho \Delta S_{t-1} + (1-\lambda)\rho \cdot \lambda \sum_{j=0}^{\infty} (1-\lambda)^j \rho^{j+1} \Delta S_{t-2-j} \tag{15}$$

Consider now (14) lagged by one period,

$$\pi_{t-1} - \frac{\alpha\lambda}{1-\lambda} y_{t-1} = \lambda \sum_{j=0}^{\infty} (1-\lambda)^j \rho^{j+1} \Delta S_{t-2-j} \tag{16}$$

Plugging (16) into (15), I obtain:

$$\pi_t - \frac{\alpha\lambda}{1-\lambda} y_t = \lambda \rho \Delta S_{t-1} + (1-\lambda)\rho \left(\pi_{t-1} - \frac{\alpha\lambda}{1-\lambda} y_{t-1}\right)$$

Finally, after some algebra, the previous equation can be written as:

$$\pi_t = \beta_1 \pi_{t-1} + \beta_2 y_t + \beta_3 \Delta y_t + \beta_4 \Delta y_{t-1}$$

which is the equation (5) used in the text. $\beta_1, \beta_2, \beta_3, \beta_4$ are the reduced form coefficients functions of the structural parameters α, λ and ρ.

B Proof of Lemma 1

I show here how to write the Sticky Information Phillips Curve,

$$\pi_t = \frac{\alpha\lambda}{1-\lambda} y_t + \lambda \sum_{j=0}^{\infty} (1-\lambda)^j E_{t-1-j} [\pi_t + \alpha \Delta y_t] \tag{17}$$

as function of the exogenous shocks. Using the standard notation, define the error of a forecast made j periods ago as:

$$\varepsilon^F_{t|t-j} = Z_t - E[Z_t \mid \Omega_{t-j}] \tag{18}$$

Use this definition to substitute out the expectations in equation (17) and obtain:

$$\pi_t = \frac{\alpha\lambda}{1-\lambda} y_t + \lambda \sum_{j=0}^{\infty} (1-\lambda)^j \delta \left(Z_t - \varepsilon^F_{t|t-j-1} \right) \tag{19}$$

where I used the linear property of the expectations. δ is a $(1 \times n)$ row vector that picks $(\pi_t + \alpha\Delta y_t)$ within the Z_t process. Equation (19) is equivalent to:

$$\frac{\alpha\lambda}{1-\lambda} y_t + \alpha\Delta y_t = \lambda \sum_{j=0}^{\infty} (1-\lambda)^j \delta \varepsilon^F_{t|t-j-1} \tag{20}$$

Let's focus now on the forecast errors in RHS of (20). In what follows I assume that the dynamics of Z_t is given by a linear system of equations.[26] In this case, each forecast error is just a linear combination of the exogenous shocks ε_t, which are defined as in the Wold decomposition of Z_t:

$$Z_t = c + \sum_{i=0}^{\infty} A_i \varepsilon_{t-i} \tag{21}$$

In details,

$$\begin{aligned} \varepsilon^F_{t|t-j-1} &= Z_t - proj\,[Z_t \mid \Omega_{t-j-1}] \\ &= c + \sum_{i=0}^{\infty} A_i \varepsilon_{t-i} - proj \left[c + \sum_{i=0}^{\infty} A_i \varepsilon_{t-i} \mid \Omega_{t-j-1} \right] \\ &= \sum_{i=0}^{j} A_i \varepsilon_{t-i} \end{aligned} \tag{22}$$

where the second equality uses (21), and the third one uses the fact that all future shocks have zero expected value.

[26] Such assumption has been the standard practice in the literature for ages. It's based on the idea that a first order linear approximation is accurate enough to describe the macroeconomic system. Nonetheless, this practice may not be completely innocuous, as noticed by Sims (2005).

Next, define $\gamma \equiv (1-\lambda)$ and use (22) to write the RHS of (20) as:

$$\begin{aligned}\lambda \sum_{j=0}^{\infty} \gamma^j \delta \varepsilon_{t|t-1-j}^F &= \lambda \sum_{j=0}^{\infty} \gamma^j \delta \sum_{i=0}^{j} A_i \varepsilon_{t-i} \\ &= \left(\delta\varepsilon_t + \gamma\delta\varepsilon_t + \gamma^2\delta\varepsilon_t + \ldots\right) + \left(\gamma\delta A_1\varepsilon_{t-1} + \gamma^2\delta A_1 \varepsilon_{t-1} + \ldots\right) + \ldots \\ &= \frac{\lambda}{1-\gamma} \sum_{i=0}^{\infty} \gamma^i \delta A_i \varepsilon_{t-i} \qquad (23)\end{aligned}$$

Finally, plugging (23) into (20) we obtain

$$\frac{\alpha\lambda}{1-\lambda} y_t + \alpha \Delta y_t = \sum_{i=0}^{\infty} (1-\lambda)^i \delta A_i \varepsilon_{t-i}$$

which proves the Lemma.

C Derivation of Orthogonality Condition (7')

From

$$\frac{\alpha\lambda}{1-\lambda} y_t + \alpha \Delta y_t = \sum_{i=0}^{\infty} (1-\lambda)^i \delta A_i \varepsilon_{t-i} \qquad (24)$$

I derived an alternative orthogonality condition, which is intended to exploit the lag structure of the variables at issue.

Write down the ARR of Z_t:

$$Z_t = \sum_{i=1}^{p} B_i Z_{t-i} + \varepsilon_t \qquad (25)$$

in its companion form

$$\widetilde{Z}_t = D\widetilde{Z}_{t-1} + \tilde{\varepsilon}_t$$

where

$$\underbrace{\widetilde{Z}_t}_{pn\times 1} = \left[\begin{array}{cccccccccc} y_t & \pi_t & X_t' & y_{t-1} & \pi_{t-1} & X_{t-1}' & \cdots & y_{t-p} & \pi_{t-p} & X_{t-p}' \end{array}\right]'$$

Now, define the $(1\times pn)$ vector $\tilde{\delta}$ as:

$$\underbrace{\tilde{\delta}}_{1\times pn} = \left[\begin{array}{cccccc} \alpha & 1 & \mathbf{O}_{1\times n-2} & -\alpha & 0 & \mathbf{O}_{1\times(n-2)(p-1)} \end{array}\right]$$

so that it picks up the argument of the expectations in (24) within the vector $\widetilde{Z}_t$. Also, define a $(1\times np)$ vector ζ as

$$\underbrace{\zeta}_{1\times pn} = \left[\begin{array}{cccccc} \frac{\alpha\lambda}{1-\lambda}+\alpha & 0 & \mathbf{O}_{1\times(n-2)} & -\alpha & 0 & \mathbf{O}_{1\times(n-2)(p-1)} \end{array}\right]$$

so that it picks the LHS of (24) within the $\widetilde{Z}_t$ process. After, write the Wold decomposition of $\widetilde{Z}_t$

$$\widetilde{Z}_t = \sum_{i=0}^{\infty} D_i \tilde{\varepsilon}_{t-i}$$

Notice that $\widetilde{Z}_t \sim VAR(1)$ so that

$$D_0 = I \ , \ D_1 = D \ , \ D_2 = D^2 \ , \ldots , \ D_n = D^n$$

Therefore, using ζ and $\tilde{\delta}$ we may write equation (6) as

$$\zeta \widetilde{Z}_t = \sum_{i=0}^{\infty} (1-\lambda)^i \, \widetilde{\delta} D^i \tilde{\varepsilon}_{t-i} \tag{26}$$

or:

$$\zeta \widetilde{Z}_t = \widetilde{\delta} \tilde{\varepsilon}_t + \sum_{i=0}^{\infty} (1-\lambda)^{i+1} \, \widetilde{\delta} D^{i+1} \tilde{\varepsilon}_{t-1-i} \tag{27}$$

I'll show that

$$\sum_{i=0}^{\infty} (1-\lambda)^{i+1} \, \widetilde{\delta} D^{i+1} \tilde{\varepsilon}_{t-1-i} \approx (1-\lambda) \, \overline{x} \cdot \zeta \widetilde{Z}_{t-1} \tag{28}$$

The terms of the summation in (28) are scalars, so

$$\sum_{i=0}^{\infty} (1-\lambda)^{i+1} \, \widetilde{\delta} D^{i+1} \tilde{\varepsilon}_{t-1-i} = tr \left(\sum_{i=0}^{\infty} \gamma^{i+1} \widetilde{\delta} D^{i+1} \tilde{\varepsilon}_{t-1-i} \right)$$

where I defined $\gamma \equiv (1-\lambda)$ to save space. Using the properties of the trace, it holds

$$\begin{aligned} tr \left(\sum_{i=0}^{\infty} \gamma^{i+1} \widetilde{\delta} D^{i+1} \tilde{\varepsilon}_{t-1-i} \right) &= \gamma \sum_{i=0}^{\infty} \gamma^i tr \left(D^{i+1} \tilde{\varepsilon}_{t-1-i} \widetilde{\delta} \right) \\ &= \gamma \sum_{i=0}^{\infty} \gamma^i x_i tr \left(D^i \tilde{\varepsilon}_{t-1-i} \widetilde{\delta} \right) \\ &\simeq \gamma \overline{x} \cdot tr \left(\sum_{i=0}^{\infty} \gamma^i \delta D^i \tilde{\varepsilon}_{t-1-i} \right) \end{aligned} \tag{29}$$

where the third equality does not hold exact. I simulate the behavior of

$$x_i = \frac{\widetilde{\delta} \widehat{D}^{i+1} \widehat{\tilde{\varepsilon}}_{t-1-i}}{\widetilde{\delta} \widehat{D}^{i} \widehat{\tilde{\varepsilon}}_{t-1-i}}$$

in the our data sample. I found that

$$x_i \simeq \overline{x} \quad \surd i$$

holds with an acceptable order of approximation. Thus, I can plug the approximated equivalence (29) in (27). Plus, from (26) lagged we have

$$tr\left(\sum_{i=0}^{\infty}\gamma^{i}\delta D^{i}\tilde{\varepsilon}_{t-1-i}\right)=tr\left(\zeta\widetilde{Z}_{t-1}\right) \tag{30}$$

So, plugging (30) into (29), and the result in (27) we obtain

$$\zeta\widetilde{Z}_{t}-(1-\lambda)\,\overline{x}\cdot\zeta\widetilde{Z}_{t-1}\approx\widetilde{\delta}\tilde{\varepsilon}_{t} \tag{9'}$$

The moment for the GMM estimation in this case is

$$E\left[\left(\frac{\alpha\lambda}{1-\lambda}y_{t}+\alpha\Delta y_{t}-\lambda\overline{x}\alpha y_{t-1}-(1-\lambda)\,\overline{x}\alpha\Delta y_{t-1}+\alpha\widehat{\varepsilon}_{t-1}^{y}\right)\cdot\mathbf{z}_{t}\mid\Omega_{t}\right]\approx0 \tag{7'}$$

where I replaced ε_{t-1}^{y} with the estimate $\widehat{\varepsilon}_{t-1}^{y}$. If the VAR (25) is correctly estimated, the GMM estimation of λ will be consistent either case.