The Role of Sticky Information in Inflation Dynamics: Estimates and Findings

Benedetto Molinari*
Universitat Pompeu Fabra†

First draft July 2004, this draft November 2006

Abstract

I derive a closed form solution for Sticky Information Phillips Curve, and I estimate it using US post-war quarterly data. My estimates of sticky information parameter are significantly lower than the calibration used by Mankiw and Reis (2002), suggesting that sticky information hypothesis is not quantitatively relevant to explain inflation persistence. The econometric strategy I pursue improves the other empirical analysis of Sticky Information Phillips Curve because it test the model with more power against the alternative model of sticky prices.

Also, I perform various tests of structural breaks to show how information stickiness has changed during the sample. This last evidence match the observed change in inflation persistence during the disinflation period in 1990's, which is an empirical conundrum for the New Keynesian framework.

Finally, I show how sticky information implies time-varying volatility of inflation. In SIPC model the variation in time of inflation volatility observed during the Great Moderation is in part endogenous, instead of being an evidence of stochastic volatility, as suggested by Cogley and Sargent (2003).

JEL Classification: E31, C51

Keywords: Sticky Information, Inflation Dynamics, 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. Contact email: benedetto.molinari@upf.edu

†Departament of Economics and Business, Universitat Pompeu Fabra (Barcelona, Spain – EU).

1 Introduction

In last years a branch of literature focused on *Rational Inattention*[1] hypothesis to explain inflation dynamics. This literature is related with early papers on limited information hypothesis of Lucas (1973), Fischer (1977), Taylor (1980), and Woodford (2001). Mankiw and Reis (2002, henceforth MR) combine some elements of Fisher and Lucas contributions and propose the Sticky Information Phillips Curve (henceforth SIPC) as theory to explain the dynamics of inflation. Their main goal was to understand why actual inflation responds gradually to shocks, as observed in post-war US economy.

MR model is based on the idea that firms absorb only sporadically the information they need for pricing their goods. When the inflow of information is limited or absent, then firms set prices based on outdated information. This idea is appealing: we could observe a failure of rational expectations equilibrium in real economies not because firms follow suboptimal pricing strategies, like adaptive rules (indexation or rules of thumb), or because they are inactive in some periods, like in Calvo (1983) framework, but because they price (rationally) their goods with imperfect knowledge. For this reason, MR presented the SIPC as an alternative theory to the New Keynesian Phillips Curve, which is based on Calvo (1983) framework.

In conclusion, MR model turns out to be an intuitive model to investigate the issue of bounded rationality in the aggregate. Yet, despite the appealing theoretical framework, there are few accompanying econometric analysis.

The objective of an empirical analysis of SIPC should be twofold. First, we are interested in testing whether the model is a valid theory to explain actual inflation dynamics. Second, it is interesting to investigate whether sticky information is actually the main source of inflation persistence, as MR claim. In facts, there is a inverse relationship between the information stickiness coefficient and the degree of inflation persistence in the model. MR show that SIPC is a better alternative to New Keynesian Phillips Curve to explain inflation persistence. After, Reis (2004) simulates SIPC and shows that fitted inflation has the same degree of persistence than actual inflation when agents update their information sets once a year. An higher coefficients, instead, implies a lower inflation persistence, and it would be consistent with a model where sticky information is not the only source of persistence. Finally, Trabandt (2003) show that both the models, i.e. SIPC and New Keynesian Phillips Curve in its Hybrid version, can trigger persistence in inflation of the same the kind of that observed in actual data.

My reason of concern with existing estimates of SIPC[2] is that their econometric strategy cannot disentangle between SIPC and Hybrid New Keynesian Phillips Curve (henceforth HNKPC). In other words, they test a specification of SIPC that fits significantly the data either when SIPC is true, or when the alternative HNKPC is true and SIPC is false. Hence, the results they find are biased

[1] The name *Rational Inattention* first appeared in Sims (1998, 2003).

[2] At the time I write this paper the only existing estimation of SIPC is Khan and Zhu (2002).

toward the acceptance of the model.[3] For this reason, I propose in this paper an econometric strategy that: (i) derives a closed form equation from SIPC that does not encompasses the HNKPC model; (ii) estimates the information stickiness parameter by GMM; (iii) analyze the effects of sticky information on inflation dynamics.

I also show how the degree of sticky information was significantly higher in the first years of my data sample than in the last ones. This findings suggests that sticky information has been an important source of inflation persistence in past periods, while today is not.

Finally, I use SIPC to analyze a different feature of inflation dynamics. Instead of focusing the analysis on inflation persistence, I argue that sticky information hypothesis may help us to understand the time varying volatility observed in actual inflation. I'll present this analysis in section 4.

The paper is organized as follows: in section 2 I review the literature on sticky information and I present MR model. Then, section 3 provides the econometric strategy and the GMM results. Section 4 is devoted to analyze the effects of sticky information on inflation dynamics. Some conclusions are given in section 5.

2 The theory

2.1 Sticky Information hypothesis

In the last decade DSGE[4] models with monopolistic competition markets and nominal rigidities (called New Keynesian models) become the standard benchmark for monetary economics. The equation that rules price index dynamics in such models is the so called New Keynesian Phillips Curve. When producers adjust their prices infrequently,[5] the inflation in model economy follows an expectations augmented Phillips Curve. For this reason the name New Keynesian Phillips Curve (henceforth NKPC).

NKPC reproduces quite well the dynamics of inflation under many aspects, but some criticism pointed out that it misses to explain inflation persistence as it is observed in actual data. In particular, it has been derived a set of main facts that NKPC fails to match: (i) actual inflation responds gradually to monetary policy shocks, while NKPC implies an immediate adjustment (Mankiw 2001); (ii) output losses typically accompany a reduction in inflation, while this is not true with NKPC (Able and Bernanke, 1998); (iii) NKPC implies that announced disinflation causes a boom, while in real economy it is the opposite (Ball 1994).

The challenge to square these facts gave the born to a number of alternative models, each based on different assumptions of firms' pricing strategy. On the one end, firms may not follow rational behavior because they don't use Rational Expectations. A well known example is the HNKPC put forward by

[3] I explain this point in section 2.5.
[4] Dynamic Stochastic General Equilibrium.
[5] Sticky prices hypothesis, Calvo (1983).

Christiano, Eichenbaum and Evans (2001): in some periods the firm does not price optimally its product, but it indexes price to the last observed change in inflation. On the other end, firms may be constrained in absorbing the relevant information they need to price optimally their products. Along this line we find those models based on inattentive agents. Regarding the reasons why information flows may be limited, the literature comes up with different explanations. In the original MR paper information constrains are exogenous. Producers update their information sets with an exogenous probability that follows a Poisson distribution, so there is no microfoundation of the model. Later on Makiw, Reis and Wolfers (2003) showed that such sticky information could explain well the main characteristic of inflation expectations surveys. So, they provide an indirect evidence to support MR assumption.

A rigorous microfoundation of SIPC based on a cost-benefit analysis of information updating is provided in Reis (2004). But alternative theories have been formulated. Information constraints may be due to physical constrains in the capacity of channels through which information is collected. Sims (1998, 2003) uses Information Theory to model the agents' inattentiveness. In this spirit, Moscarini (2004) solves profit maximization problem for producers with limited information channels, and he finds a behavioral equation for aggregate prices level and inflation dynamics; Luo and Young (2005) simulate a fully fledge DSGE model with Rational Inattention à la Sims both on the aggregate demand and supply side of the economy. They find that rational inattention does not provide any relevant improvement to the dynamics of a general equilibrium model.

It worth notice that both *Indexation* and *Inattention* models reproduce well the observed persistence in actual inflation, as showed by Trabandt (2003), and Luo and Young (2005). In particular, Trabandt (2003) simulates fully fledge DSGE models[6] using either HNKPC or SIPC as aggregate supply equations. He finds that both models match the three facts listed above, i.e. the ones that NKPC couldn't match. Moreover, he showed that inflation dynamics is persistent in both the models, as it is in actual data.

2.2 Empirical evidences of Sticky Information: Expectations Surveys

MR model is based on a reasonable and intuitive argument: what would be the outcome of a model where one solves the expectations involved in producers' pricing strategy using a rule which is able to reproduce the main stylized facts about agents' expectations, as they are reported in surveys of inflation expectations? Hereafter I present the surveys of expectations which MR linger over to support the sticky information assumption.

Using microdata from the Surveys of agents' expectations,[7] Mankiw, Reis

[6] Typically *fully fledge model* is meant a model with an equation for aggregate supply (in this case either HNKPC or SIPC), an equation for aggregate demand (the IS alike equations), and an equation for nominal interest rate adjustment (the "policy equation", for instance the Taylor rule).

[7] The two surveys in considereation in Makiw et al. (2003) are the Michigan Survey of

and Wolfers (2003) underlined that disagreement about the future level of inflation is one crucial characteristic of expectations surveys, both among Consumers and Professional forecasters. In particular, they find that interquantile range of inflation expectations for 2003 among professional economists goes from 1.5 to 2.5 percent.[8] Among the general public, the interquantile range of expected inflation goes form 0 to 5 percent,[9] see Figure 1. Very similar findings are obtained in Carroll (2001), and pioneered in Roberts (1998), (2001).

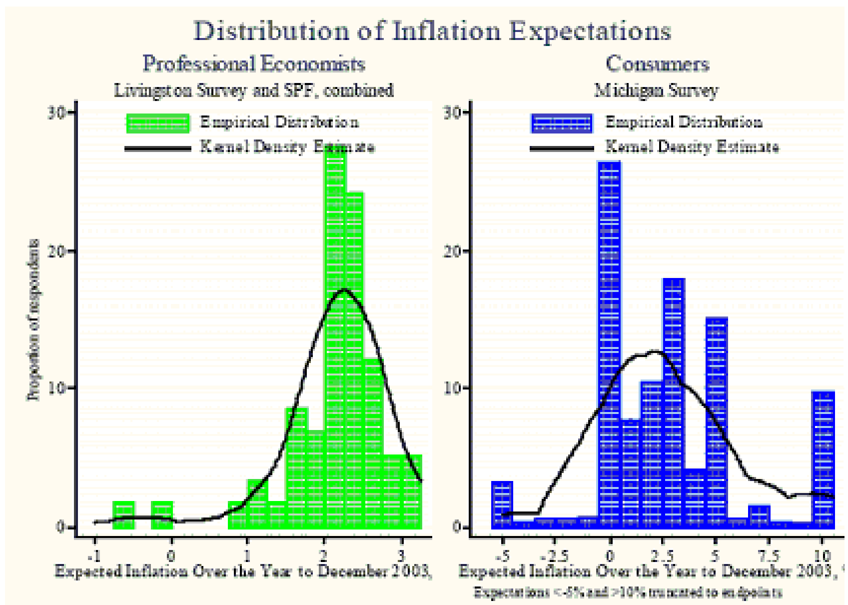

Figure 1. Cross-section distribution of Expectations (Source Mankiw et al., 2003)

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.

Consumers' Expectations, and the Survey of Professional Forecasters.

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

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

3. The forecasting errors are persistent, in the sense that realized last year prediction error has a positive correlation with today forecast error. See Figure 2. (source Kenny, 2002)

4. Not all the information available at time t is fully exploited for next periods forecasts made in t and after.[10]

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.[11]

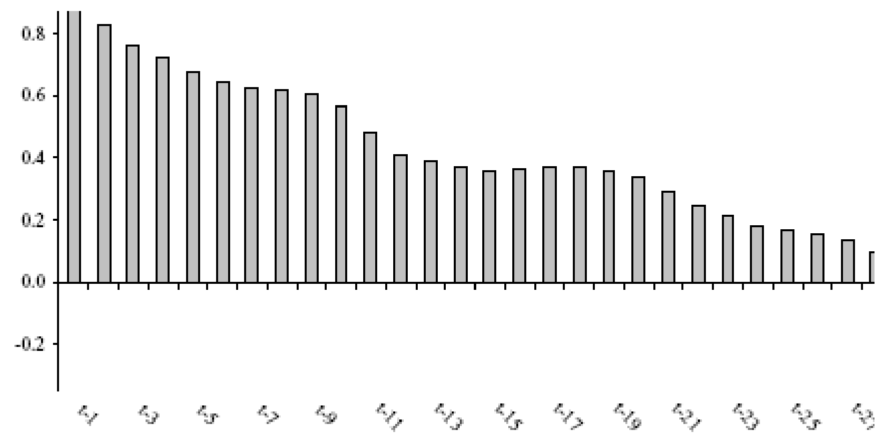

Figure 2. Expectations errors. A positive and persistent autocorrelation beyond $t-12$ indicate a violation of RE hypothesis.

2.3 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*

[10] *Available at time t* means the information accessible to agents in period t. For instance, one may think that only variables dated $t-1$ and before are available at time t.

[11] It's worth noticing that autocorrelation in forecast errors is not itself a test of rationality unless the errors are not overlapping, as argued by Batchelor (1982). Indeed, if agents didn't know the realization of errors at the time they forecast, then clearly these errors would add up in the forecast errors. 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) to check whether the RE hypothesis is violated.

RE hypothesis. Consequently, some firms will form expectations conditional on outdated information sets.

In particular, MR assumed that only a constant fraction of the population gets new information in the period, while the other firms choose their optimal actions using expectations conditional on outdated information. In other words, they decided the price for period t when they last updated the information in $t-j$, and after then they stick to their j period ago forecast about the price to be charged today.

Mankiw at al. (2003) show that such rule does a pretty good job in reproducing the facts listed in section 2.1. Plus, it explains why the expectation 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 and he solves the optimal time-dependent prices adjustment rules for firms. His model with limited acquisition of relevant information generate inertia in inflation.

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.

Finally, 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 the theory of epidemies, 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.

2.4 The Sticky Information Phillips Curve

Suppose that the economy has monopolistic competition market populated by a continuum of *ex-ante identical*[12] producers, each producing a single imperfect

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

Nonetheless, in period t when the information acquisition occurs, they are not be identical anymore, because some of them update the information, thus setting optimal prices in t, while the others will keep the price they set j periods ago, which is conditional on outdated substitute good; and assume perfect competition on factors market.

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,[13] an higher demand means that each firm is likely to raise its 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.

Starting from this framework, MR assumed that every period a fraction $\lambda \in (0, 1]$ of firms update the information about the current state of economy, and compute optimal prices conditional on such information. The rest of firms sets today price conditional on outdated information. Each firm has the same probability to update information, regardless of how long has been since its last update.

So each firm *i,* which updated the information 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 firms are ex-ante identical, 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.

From this assumption MR derived a behavioral equation for aggregate price level in the economy, where "short-run Phillips curve is apparent... output is positively associated with surprise movements in the price level".[14] 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 in the model evolves accordingly to:

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

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.[15] Thus, today's inflation

information.

[13] 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.

[14] Mankiw and Reis (2002) pg. 1300

[15] For the proof and the details see Mankiw and Reis (2002).

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 past periods expectations of today's inflation and output growth.

The mechanism is the following: in period t occurs a shock that increases output gap. Accordingly to (3), inflation raises contemporaneously because of the trade off term $\frac{\alpha\lambda}{1-\lambda}y_t$. Then, in period $t+1$ a fraction λ of agents gets aware of the shock occurred in t and inflation raises again because the term $E\left[\Delta y_{t+1} \mid \Omega_{(t+1)-1-j}\right]$ is positive for $j<1$. The same happens in period $t+2$, when a fraction $\lambda(1-\lambda)$ gets aware of the shock, and so on for all following periods, when the effect of the shock on inflation fades out at rate $(1-\lambda)^j$. Hence, the same shock affects inflation level for infinite periods, and the inflation in the model turns out to be serially correlated, as real data suggest.

2.5 Estimates of Sticky Information Phillips Curve

As I said, the main purpose of SIPC was to solve the counterfactual implications of NKPC (see page 3). In their paper, MR (2002) simulate SIPC and they show how it responds gradually to Monetary Policy shocks, and how the inflation in the model has an highly persistent time path. Thus, MR claim that SIPC improves the ability of NKPC to reproduce actual inflation dynamics.

However, MR simulate the model using only one particular calibration of information stickiness parameter, i.e. $\lambda = 0.25$. This value means that agents update information in average once a year, and MR argue that this is a reasonable value because it is in accordance with the empirical evidences coming from surveys of agents' inflation expectations.[16] Regarding this, I will just mention that it can be misleading to trust micro data to calibrate aggregate models because of the aggregation issues embedded in macroeconomic models.[17] Therefore, a proper estimation of SIPC seems necessary to assess the validity of the model. Surprisingly, as far as I know, there is only one estimation of SIPC, by Khan and Zhu (2002).

Khan and Zhu estimate SIPC for U.S., Canada and UK, finding values of lambda around 0.3, so almost in line with the calibration used by MR. Their econometric strategy is the following: first, they truncate the infinite sum of expectations in equation (3) at $t-j_{\max}$; then, they substitute the remaining expectations terms with predictions of a VAR model set ad-hoc to forecast inflation and output gap. For instance, the expectation about today inflation and output gap growth conditional on information dated $t-5$ is replaced by $proj\left[\pi_t+\alpha\Delta y_t \mid z_{t-5}, z_{t-6}, ...z_{t-(5+p)}\right]$ where z_t is a vector of macroeconomic variables and p is the order lag of the VAR model.

[16] See Mankiw, Reis, and Wolfers (2003)

[17] One well known example is the case of Fisher (labor supply) elasticity. Standard Real Business Cycle model fits the actual aggregate labor volatility usingn a value of Fisher elasticity close to 1, whereas the Fisher elasticity that comes out from surveys of workers is around 1/6.

The estimation strategy of Khan and Zhu is criticizable for two reasons. The first one is technical: they use stochastic regressors without adjusting the standard errors of estimated coefficients to account for it. Therefore the confidence intervals are bigger than what they report, and we cannot rely on their test of hypothesis. In particular, we don't know whether MR calibration of λ is accepted or rejected by the data.

Second, Khan and Zhu basically exploit persistence in inflation to achieve an estimation of sticky information coefficient λ. In fact, when they replace the expectations in equation (3) with the VAR predictions, they are proxying the expectations terms with a linear combination of past inflations and output gaps (plus past values of the other variables included in the vector z_t). So, they regress inflation on $t - j_{\max}$ lags of inflation and output gap. Since such model encompasses HNKPC, then their specification will be significant also if SIPC is false and the alternative HNKPC is true. Hence, Khan and Zhu provide a test for SIPC with small power against its main competitor theory.

Moreover, if actual inflation come from a model with heterogeneous agents, where some of them use indexation, some others are inattentive, and some others are rational, then Khan and Zhu's estimation of λ is downward biased, i.e. it detects an higher level of sticky information among the firms). That's because their strategy addresses all the persistence of inflation to the unique source of sticky information.

2.6 Why not to estimate SIPC using inflation level

I argued above that a direct estimation of equation (3) is not a powerful test for SIPC against the alternative model of HNKPC. The intuition is clear if we observe the two models. On one side, Hybrid New Keynesian Phillips Curve is given by

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

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

On the other side, under some reasonable assumptions, SIPC can be written as:[18]

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

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

Now, suppose one estimates equation (5) for US, but the true model is (4). Since in US data output gap leads inflation,[19] then the coefficient β_3 will be significant because π_{t+1} is a missing variable in the specification (5) and Δy_t works as proxy. Also β_4 will be significant, because of the correlation between Δy_{t-1} and π_t in actual data. In the end, λ will be a significant parameter even though the SIPC model is false!

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

[19] Galì and Gertler (1999) raise the point. When we observe an increase in output at period t, i.e. positive output growth in t, then we expect increase in following period inflation.

Thus, if we estimate SIPC using the structural equation (3) we will find significant coefficients even when the SIPC model is false and the alternative HNKPC model is true, i.e. we don't control for type II error.

3 The Estimation

Instead of estimating directly equation (3), as other papers do, I propose a different approach in order to estimate SIPC.

First, I write SIPC in terms of exogenous errors, so I get rid of the expectations terms in (3); this result is provided hereafter as Lemma 1. Second, using Lemma 1, I find the variance covariance (hereafter VCV) matrix of exogenous errors as functions of SIPC parameters. It is worth noticing that the VCV equations I compute have a finite number of terms. My first contribution is methodological: I show how to avoid the infinite dimension problem usually associated with estimations of SIPC without using any truncation or approximation.

Third, I estimate a set of vector autoregressions (VAR) to obtain the exogenous errors for inflation and output gap in the sample period. These shocks are the linear counterpart to RE errors. Then, I use them to compute the VCV matrix described above. Fourth, I use the VCV structural equation I derived as orthogonality conditions to estimate the information stickiness parameter λ with GMM.

This approach exploits a structural equation involving the VCV matrix of errors that is true only when SIPC holds. In particular, my specification does not encompass the HNKPC model. Hence, I provide an econometric strategy that disentangle between the two theories, which implies a more powerful test for SIPC against the alternative model of HNKPC.

3.1 Econometric specification

As in standard macroeconomic dynamics models, I assume here that inflation and output gap dynamics result from the interaction of n macroeconomic variables, which I define here as elements of a covariance-stationary vector process Z_t. This assumption poses very few structure on inflation and output gap processes, nonetheless it allows me to find a useful result:

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 SIPC (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 a $(n \times 1)$ vector of exogenous shocks. δ is a $(1 \times n)$ row vector that picks up $(\pi_t + \alpha \Delta y_t)$ within Z_t.

Proof. See Appendix B. ■

Now, multiply equation (6) by $(\delta\varepsilon_t)'$ and take the expectations at time t, then:

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

where $\Sigma \equiv E\left[\varepsilon_t\varepsilon_t'\right]$ is the VCV matrix of errors.

The orthogonality condition (7) relates the VCV matrix of shocks with the SIPC parameters α, λ and it is handy for estimations because it works out the infinite summation in (3).

One difficulty with equation (7) is that it is function of ε_t and Σ, which are unknown. However, it can be shown that if $\widehat{\varepsilon}_t^{\Delta y}, \widehat{\varepsilon}_t^{\pi}, \widehat{\Sigma}_T$ are consistent estimators of $\varepsilon_t^{\Delta y}, \varepsilon_t^{\pi}, \Sigma$, then the sample analog

$$\sum_{t=1}^{T}\left[\left(\frac{\alpha\lambda}{1-\lambda}y_t + \alpha\Delta y_t\right)\left(\alpha\hat{\varepsilon}_t^{\Delta y} + \hat{\varepsilon}_t^{\pi}\right) - \delta\widehat{\Sigma}_T\delta'\right] = 0 \tag{8}$$

converges almost surely to the population moment (7).

Hence, I estimate (8) where the candidate regressors – $\widehat{\varepsilon}_t^{\Delta y}$, $\widehat{\varepsilon}_t^{\pi}$ and $\widehat{\Sigma}_T$ – come from a $VAR(p)$, which includes inflation and output gap, and where p is the minimum order of lags to have no autocorrelation in residuals.

3.2 VAR estimation

As I said, my strategy consists in finding consistent estimators for $\varepsilon_t^{\Delta y}, \varepsilon_t^{\pi}, \Sigma$. The candidates are $\{\varepsilon_t(\beta), \Sigma(\beta)\}\ |_{\beta=\beta_T^{VAR}}$ In particular, I estimate various VAR models using different sets of variables. I provide two main specifications: the baseline, which includes inflation, output gap and interest rate. A second one that I name *min RMSE,* which includes all the variables that improve in a relevant way the forecast of inflation and output gap according to Stock and Watson (2003).[20]

In both specifications inflation is measured either as the CPI, or as the implicit GDP deflator; and output gap is detrended either with HP, or BP filter. Considering all the possible permutations, I have 8 VAR specifications. The complete set of variables in the *min RMSE* specification includes: short-term

[20] 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.

In details, I estimate a VAR model with a vector of endogenous variables Z_t. This vector is,

$$\underbrace{Z_t}_{n\times 1} = \left[\begin{array}{ccc} \Delta y_t & \pi_t & X_t' \end{array}\right]'$$

where X_t can be either $X_t = i_t$ (the baseline specification), or a $(n-2\times 1)$ vector which includes all the variables that bring relevant information to minimize the adjusted $RMSE$ of output gap and inflation (the *min RMSE* specification).

interest rate (the Fed Fund Rate); either the long-term interest rate (10 years Gov't bond), or the term spread (long run minus short run interest rate); the real Stocks Price Index (S&P500, deflated by CPI); the IMF price index of commodities; the real money (real M2 minus small time deposits); unemployment rate; Total Capacity Utilization rate (TCU).

The sample goes from 1957q1 to 2005q4; all the variables are taken in logs except for unemployment, total capacity utilization, and interest rates rates. Also, the variables are detrended or differenced when necessary. All the series used in VAR estimation are stationary.

Finally, every VAR specification has the minimum number of lags in order for the residuals to be not serially correlated. Autocorrelation of residuals is tested with a standard LM test.

All the variables, both for GMM and VAR estimations, come from FRED II database of US economy.[21]

3.3 GMM Results

First, I multiply equation (8) by a vector of instruments x_t, which contains all variables dated $t-1$ and before.[22] Assuming i.i.d. errors ε_t I obtain

$$\sum_{t=1}^{T}\left[\left(\left(\frac{\alpha\lambda}{1-\lambda}y_t+\alpha\Delta y_t\right)\left(\alpha\hat{\varepsilon}_t^{\Delta y}+\hat{\varepsilon}_t^{\pi}\right)-\delta\widehat{\Sigma}_T\delta'\right)\cdot x_t\right]=0 \tag{9}$$

Next, I estimate the system of equations (9) with GMM using $\{\varepsilon_t(\beta),\Sigma_T(\beta)\}\mid_{\beta=\beta_T^{VAR}}$ as regressors.[23] The sample goes from 1958q4 to 2005q4 (189 obs.). The GMM sample is shorter than the one used in VAR estimations because I loose 7 lags to obtain VAR estimates.

To control for small sample bias in nonlinear GMM estimator, I estimate two alternative specifications of (9). The first one is (9) times $(1-\lambda)$; the second one is (9) times $\frac{1-\lambda}{\alpha\lambda}$. They are referred to as (1) and (2) in next tables.

Estimation results are summarized in table 1.

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

[22] The set of instruments contains: a constant, 4 lags of inflation, 4 lags of output gap, two lags of unemployment rate, interest rate, marginal cost, money growth, and the term spread.

[23] In this case the GMM estimator is the Non-linear IV estimator. A choose the optimal weighting matrix to minimize the variance of the moments.

Restricted α = .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)

Table 1. 2-step GMM with optimal weighting matrix. US data, sample 1958q4 – 2005q4. HP filter for output gap. Newey-West HAC standard errors adjusted for stochastic regressors. p-values in parenthesis. J statistic is the Hansen test of overidentifying restrictions (19 moments).

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.[24]

I estimate only parameter λ, while I calibrate the α. I don't estimate α because I don't use any relevant data to pin it down. Specifically, in MR model parameter α depends on intertemporal elasticity of substitution of consumers, on Fisher elasticity of labor, and on the elasticity of demand for single-variety goods. Since in the estimation of (9) I don't use data on consumption, nor on labor, and I don't estimate any aggregate demand equation, then I didn't attempt to estimate α.

The estimates of information stickiness parameter λ are our main concern. In all specifications λ_T^{Gmm} is in the range assumed by the theory, i.e. within the $(0, 1]$ interval. More precisely, it ranges between $[0.75, 0.95]$. This value implies that firms update the information on average three times per year, i.e. every quarter and a month. As a matter of fact, this value is far away from MR and Reis (2004) calibration (once per year). This finding is not surprising: Reis (2004) showed that MR calibration of $\lambda = 0.25$ is the correct value for SIPC to replicate the persistence of actual inflation, *as if* sticky information was only source of persistence. Yet, I explained above how misleading is such exercise given that actual inflation may come from a model where some agents are inattentive and some others use adaptive rules. On the contrary, the approach I pursue here should pin down the exact degree of sticky information no matter how much it contributes to inflation persistence.

[24]Standard distributions for hypothesis testing with IV estimators are reliable only if instruments are not weak.

Unfortunately, is still unclear how to check for weak instruments in GMM estimations with nonlinear orthogonality conditions and possibly nonspherical residuals. Therefore I'll test the hypotheses using standard distributions.

However, I assess MR calibration testing the null hypothesis $\lambda = 0.25$. In all 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. Since MR model encompasses RE case when $\lambda = 1$, then my second concern has been to test whether RE hypothesis is accepted or rejected by the data. The null is rejected at 5% level in all the $\{\Delta y_t, \pi_t, i_t\}$ specifications, but it is accepted in the *min RMSE* ones. So, for RE hypothesis my results are not overwhelming in any direction.

3.4 Robustness analysis

3.4.1 Empirical robustness

I checked the robustness of results along four dimensions:

1. Sensitivity to calibration of α;
2. Different assumptions about Z_t process. For instance, I assumed that $Z_t \sim AR(2)$, then I solved equation (7) under this hypothesis, and I estimate the resulting equation.
3. A different filter for gap variables, i.e. Band Pass (BP) instead of Hodrick Prescott (HP);
4. A different orthogonality condition, (7'), derived from (6) (see Appendix C.)

The following Table 2. summarize the results.

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)
$Z_t \sim AR(2)$	.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)

Table 2. 2-step GMM with optimal weighting matrix. US data, sample 1958q4 – 2005q4. HP filter for output gap except than in the 3rd case where I use Band-pass (6,32) filter. Newey-West HAC standard errors adjusted for stochastic regressors. p-values in parenthesis. J statistic is the Hansen test of overidentifying restrictions (19 moments).

From robustness analysis we can draw the following conclusions:

- Results from baseline estimation are basically confirmed (cfr. table 1 with table 2). The null hypothesis of $\lambda = 0.25$ is always rejected. So, MR calibration fails to fit the actual data in all the estimations. Thus, the main conclusions drawn from estimations in section 3.1 are confirmed.
- In general it turns out that the smaller are forecast errors $\{\widehat{\varepsilon}_t\}_{t=1}^{T}$, the bigger is λ_T^{Gmm}.
- It uncovers a stable *inverse* relationship between α and λ.

3.4.2 Theoretical robustness

I derive a different orthogonality condition from equation (6), to check if results shown in table 1. are robust to changes in orthogonality conditions. In detail, I square equation (6), and I take the expectations; i.e.

$$E\left[\frac{\alpha\lambda}{1-\lambda}y_t + \alpha\Delta y_t\right]^2 = \sum_{i=0}^{\infty}(1-\lambda)^{2i}\,\delta A_i \Sigma A_i' \delta' \tag{10}$$

The main difference using (7) or (10) to estimate λ is apparent: equation (7) exploits the VCV matrix of errors Σ as function of λ, while equation (10) exploits the effect of sticky information λ on the variance of a linear combination of output gap and output gap growth. Consequently, λ disappears from RHS of (7), and the estimation only exploits the information about λ contained in LHS of (6). In (10), instead, λ remains as a coefficient of the summation, which determines its rate of convergence. So, if we estimate λ using (10) we exploit the information contained both in RHS and LHS of (6).

To estimate equation (10), first I compute the sample mean of LHS (10) using actual data. Then, I simulate the RHS of (10) using the VCV matrix of errors $\widehat{\Sigma}$, and the dynamic multipliers $\hat{A}_i$ coming from the VAR (see section 3.2).

It is worth noticing that a solution of equality (10) for $\lambda \in (0,1)$ must always exist for some values of α. Indeed, on the one end of the $(0,1)$ interval, i.e. for $\lambda \rightarrow 0$, we have that LHS of (10) goes to a quantity that depends on α, i.e. $E\left[\frac{\alpha\lambda}{1-\lambda}y_t + \alpha\Delta y_t\right]^2 \rightarrow \alpha^2 E(\Delta y_t^2)$, whereas the RHS of (10) 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, i.e. for $\lambda \rightarrow 1$, the LHS of (10) goes to plus infinity because of the first term, whereas the infinite summation in RHS of (10) collapses to its first term, $\delta\Sigma\delta' < \infty$; thus, when λ increases, at some point it will hold $LHS > RHS$ for any value of alpha chosen. Therefore, for any $\alpha \leq \bar{\alpha}$ there exists at least one $\lambda \in (0,1)$ s.t. $LHS = RHS$.

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

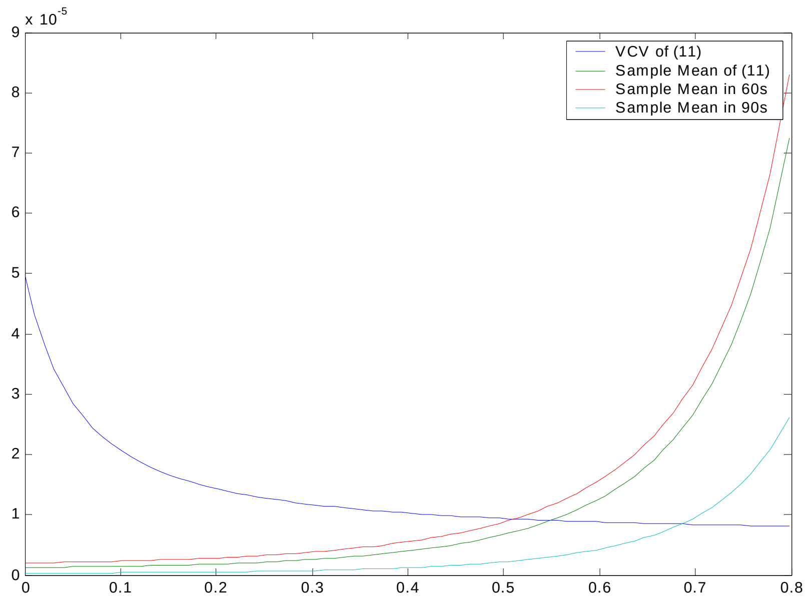

Figure 3.

As we expected there is a point of intersection, but the estimate of λ here differs substantially from the previous ones. With the calibration of α used above, (10) holds for $\lambda \approx .55$. 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

In previous section I argued that sticky information is not a relevant issue to explain inflation persistence. At this regard, two questions arise: (i) is this conclusion true overall the sample? or maybe in some periods sticky information was a main source of inflation persistence? (ii) is there something else we can learn about inflation dynamics from the analysis of sticky information?

To answer to the first question, I check whether the coefficient λ has had structural breaks during the considered period. Actually, I'll show that in 60's the degree of information stickiness was higher than later on. Accordingly, sticky information was a more important source of inflation persistence 40 years ago than today.

The second issue regards the volatility of inflation. In SIPC the information stickiness parameter λ affects not only the persistence, but also the variance of inflation. Interestingly, this effect of λ on the variance is not monotone. In a model with sticky information for a given variance of the exogenous shocks that affect output gap and inflation, if the sign of covariance between the shocks is positive the variance of inflation will be bigger than if the covariance is negative.

This characteristic is typical of sticky information model. In HNKPC the parameter that controls for inflation persistence, i.e. the backwardness parameter, has a monotone (positive) effect on inflation variance. To say, if the backwardness parameter increases then both persistence and variance increases, no matter the covariance among exogenous shocks. While in SIPC if λ increases, then the variance may increase or decrease depending on the covariance between the exogenous shocks.

In section 4.2 I argue that this typical effect of sticky information may explain part of the reduction of inflation volatility observed in 90's.

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 why disinflation in 90's was accompanied by fall in inflation persistence.[25] 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..."[26]

In SIPC inflation persistence depends on firms knowledge of economic indicators. It makes sense to think that in recent years firms have had a better knowledge than in the past. Therefore, the information stickiness should have diminished, i.e. λ increased, during the disinflationary period in 90's. Since the econometric strategy used in this paper matches the contemporaneous VCV matrix of exogenous shocks to actual data, then λ is estimated without any link to inflation persistence. Thus, in principle there is no reason why we should observe a structural change in λ to match the change in inflation persistence. This makes the test not trivial.

I perform Andrews (1993) supLM test of structural breaks.[27] This test cuts the tails of the sample and computes the most likely point in time where a break might have occurred in the middle subsample. I summarized the results in table 3. I run the test for all the specifications tried above, and I report range, mean, and median of the supLM statistics.

[25] The point was made in Erceg and Levin (2003) and Bayoumi and Sgerri (2004).

[26] Bayoumi and Sgerri (2004) pg. 6

[27] I choose Andrews's supLM because it is the most powerful test when the timing of the (possible) break in the series is unknown.

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

Table 3. π_0 indicates the percent of each tail cut. The supLM test has a non standard distribution. The asymptotic critical values are given in Andrews (1993). *, **, *** means significance respectively at 1%, 5%, and 10%

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 ($\pi_0 = 0.10$). 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 1970's the increase in inflation volatility biases the results and the test captures the spurious effect of oil shock as structural break in λ – see Figure (4.)

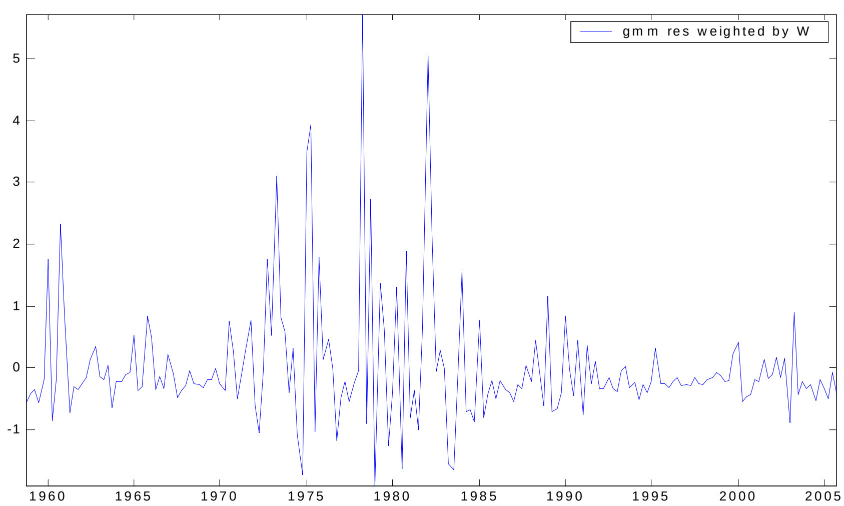

Figure 4.

Therefore, I perform a second set of tests to control for exogenous volatility bias in 1970's. Under the null hypothesis that the same model holds throughout the sample, I test whether λ estimates are equal in two subsample: one goes from 1959q1 to 1970q1, and the second from 1990q4 to 2005q4. I compare the coefficients using LM and Wald tests. Results are in table 3. above.

The null hypothesis $\lambda_{60}^{Gmm} = \lambda_{90}^{Gmm}$ is now rejected almost in all specifications. In other words, it is very likely that a structural break occurred between the first and the last years of the sample. It make sense: 40 years ago information is likely to have been stickier. Less media to channel macroeconomic informations, less accurate forecasts and previsions, less experienced authorities, less data gathering, etc. It is not surprising that firms acquired information slower, therefore taking longer time to react to new events.

Accordingly with this finding, while sticky information was an important source of inflation persistence in the past, nowadays this is not true anymore, and we can explain part of the reduction of inflation persistence through this channel.

4.2 Volatility Analysis

To see the effect of sticky information on inflation volatility, it is convenient to write equation (7) as:

$$\sigma_{\pi}^{2} = \frac{\alpha\lambda}{1-\lambda} cov(y_t, \varepsilon_t^{\pi}) - \alpha\sigma_{\pi,\Delta y} + \frac{\alpha^2\lambda}{1-\lambda} cov(y_t, \varepsilon_t^{\Delta y}) \tag{11}$$

where σ_{π}^{2} is the variance of inflation. To obtain (11) I applied the expectation operator to each term, and I solved the matricial product $\delta\Sigma\delta'$.

Now, taking the derivative of σ_{π}^{2} w.r.t. λ we have,

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

So, an increase of information stickiness, i.e. a reduction of λ, increases the variance of inflation only if the covariance between the shocks is negative and in absolute value bigger than α; i.e.

$$\begin{aligned} & \frac{\partial\sigma_{\pi}^{2}}{\partial\lambda} \leq 0 \\ iff \quad & -\frac{cov(y_t, \varepsilon_t^{\pi})}{cov(y_t, \varepsilon_t^{\Delta y})} \geq \alpha \end{aligned} \tag{12}$$

Suppose that in one period the exogenous shocks are such that (12) holds, and in the following period they have same variances but opposite covariance. In this case there will be a reduction in the variance of inflation.

Observe now what happened during the 1990's. In Figure (5.) I compute a rolling window of annual means of (12) (the blue line). The green line is $\alpha = 0.2$

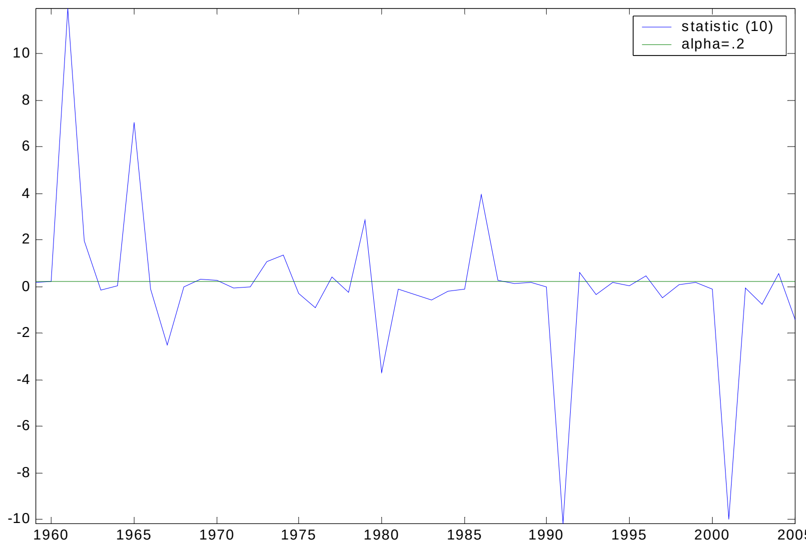

Figure 5.

The picture shows that during Great Moderation[28] the statistics (12) was repeatedly negative, implying that sticky information operated in reducing inflation variance.

This finding can be interpreted as one of the causes of the inflation volatility reduction observed during the Great Moderation. In this perspective such reduction cannot be addressed entirely to some stochastic reduction of exogenous shocks variances, as claimed by Cogley and Sargent (2003), because part of it can be explained through sticky information channel.

5 Conclusions

I presented results where SIPC theory is validated by US post war data, but information stickiness seems not to be a crucial issue to explain inflation persistence. My estimates of sticky information parameter, $\lambda_T^{Gmm} \simeq 0.8$, points an information updating around 3 times per year, while Reis (2004) showed that simulated inflation from SIPC has the same degree of persistence of actual inflation if firms update information once per year (i.e. $\lambda = 0.25$). Moreover, the estimates of λ_T^{Gmm} are more sensible to the accuracy of forecasts[29] than to the chained structure of forecast errors, which is the trademark of sticky information model. This evidence is more in line with a model where the key issue

[28] 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).

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

is the agents' uncertainty about the true model to be used for forecasts rather than sticky information, as suggested by Branch (2004).

All in all, my analysis of SIPC suggests that sticky information cannot be a ultimate theory of inflation dynamics, and we cannot dismiss yet those theories that rely on adaptive producers' behavior in order to explain inflation persistence.

On the other side, SIPC does a good job in explaining the time varying volatility of inflation. In particular, part of the reduction in inflation volatility observed during the Great Moderation in 90's can be explained with the different covariance of the exogenous shocks in this periods with respect to previous years. This covariance, indeed, affects the level of inflation variance in SIPC model.

A conclude saying that an interesting line of future research would be to give a quantitative assessment of this last finding. In particular, this can be done with the estimation of a fully fledge model where sticky information theory is used for the identification of exogenous shocks.

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] Calvo, Guillermo (1983): 'Staggered prices in a utility-maximizing framework', Journal of Monetary Economics n^o 12, pp. 383-398

[7] 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

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

[9] Christiano J. Lawrence, and Eichenbaum Martin, and Evans Charles (2001): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy", NBER wp 8403

[10] 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

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

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

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

[14] Kenny, Geoff and Forsells, Michael (2002): "The Rationality of Consumers Expectations: Survey-based Evidence for Euro area", European Central Bank wp 163

[15] 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

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

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

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

[19] Mankiw Gregory N. 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

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

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

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

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

[24] ———————— (2003): "Implications of rational inattention", Journal of Monetary Economics n^o 50, pp. 665-690

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

[26] ——————————————- (2003): "Has the Business Cycle Changed? Evidence and Explanations", forthcoming FRB Kansas City symposium, Jackson Hole, Wyoming, August 2003

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

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

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

[30] ——————— (2003): "Interest and Prices", Princeton University Press

A Derivation of Equation (5)

Define

$$S_t = p_t + \alpha y_t \tag{13}$$

and assume that growth rate of S_t to follow an AR(1) process, i.e.

$$\Delta S_t = \rho \Delta S_{t-1} + u_t \tag{14}$$

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

Now, plugging (14) in SIPC (3) I obtain:

$$\pi_t - \frac{\alpha\lambda}{1-\lambda} y_t = \lambda \sum_{j=0}^{\infty} (1-\lambda)^j E_{t-1-j}\left[\Delta S_t\right] \tag{15}$$

Now, given (14), it is $E_{t-1}\left[\Delta S_t\right] = \rho \Delta S_{t-1}$, $E_{t-2}\left[\Delta S_t\right] = \rho^2 \Delta S_{t-2}$, and so on. Thus, I work out expectations in (15) and I have:

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

or,

$$\pi_t - \frac{\alpha\lambda}{1-\lambda} y_t = \lambda\rho\Delta S_{t-1} + (1-\lambda)\rho \cdot \lambda \sum_{j=0}^{\infty} (1-\lambda)^j \rho^{j+1} \Delta S_{t-2-j} \tag{17}$$

Now, consider (16) lagged by one period,

$$\pi_{t-1} - \frac{\alpha\lambda}{1-\lambda} y_{t-1} = \lambda \sum_{j=0}^{\infty} (1-\lambda)^j \rho^{j+1} \Delta S_{t-2-j} \tag{18}$$

and use (18) into (17) to get rid of the infinite sum:

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

After some algebra, 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,

$$\pi_t = \frac{\alpha\lambda}{1-\lambda} y_t + \lambda \sum_{j=0}^{\infty} (1-\lambda)^j E_{t-1-j} [\pi_t + \alpha \Delta y_t] \tag{19}$$

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

$$\varepsilon^F_{t|t-j} = Z_t - E[Z_t \mid \Omega_{t-j}] \tag{20}$$

Use this definition to substitute out the expectations in equation (19) and obtain:

$$\pi_t = \frac{\alpha\lambda}{1-\lambda} y_t + \lambda \sum_{j=0}^{\infty} (1-\lambda)^j \delta \left(Z_t - \varepsilon^F_{t|t-j-1} \right) \tag{21}$$

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 (21) is equivalent to:

$$\frac{\alpha\lambda}{1-\lambda} y_t + \alpha \Delta y_t = \lambda \sum_{j=0}^{\infty} (1-\lambda)^j \delta \varepsilon^F_{t|t-j-1} \tag{22}$$

Let's focus now on the forecast errors in RHS of (22). In what follows I assume that the dynamics of Z_t is given by a linear system of equations.[30] 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:

$$Z_t = c + \sum_{i=0}^{\infty} A_i \varepsilon_{t-i} \tag{23}$$

In details,

$$\begin{aligned} \varepsilon^F_{t|t-j-1} &= Z_t - proj [Z_t \mid \Omega_{t-j-1}] \\ &= c + \sum_{i=0}^{\infty} A_i \varepsilon_{t-i} - proj \left[c + \sum_{i=0}^{\infty} A_i \varepsilon_{t-i} \mid \Omega_{t-j-1} \right] \\ &= \sum_{i=0}^{j} A_i \varepsilon_{t-i} \end{aligned} \tag{24}$$

where the second equality uses (23), and the third one uses the fact that all future shocks have zero expected value.

[30] 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 (24) to write the RHS of (22) as:

$$\begin{aligned}
\lambda \sum_{j=0}^{\infty} \gamma^j \delta \varepsilon_{t|t-1-j}^F &= \lambda \sum_{j=0}^{\infty} \gamma^j \delta \sum_{i=0}^{j} A_i \varepsilon_{t-i} \\
&= \left(\delta\varepsilon_t + \gamma\delta\varepsilon_t + \gamma^2\delta\varepsilon_t + \ldots\right) + \left(\gamma\delta A_1\varepsilon_{t-1} + \gamma^2\delta A_1\varepsilon_{t-1} + \ldots\right) + \ldots \\
&= \frac{\lambda}{1-\gamma} \sum_{i=0}^{\infty} \gamma^i \delta A_i \varepsilon_{t-i} \qquad (25)
\end{aligned}$$

Finally, plugging (25) into (22) 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

$$\frac{\alpha\lambda}{1-\lambda} y_t + \alpha\Delta y_t = \sum_{i=0}^{\infty} (1-\lambda)^i \delta A_i \varepsilon_{t-i} \qquad (26)$$

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:

$$Z_t = \sum_{i=1}^{p} B_i Z_{t-i} + \varepsilon_t \qquad (27)$$

in its companion form

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

where

$$\underbrace{\widetilde{Z}_t}_{pn\times 1} = \begin{bmatrix} y_t & \pi_t & X_t' & y_{t-1} & \pi_{t-1} & X_{t-1}' & \cdots & y_{t-p} & \pi_{t-p} & X_{t-p}' \end{bmatrix}'$$

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

$$\underbrace{\tilde{\delta}}_{1\times pn} = \begin{bmatrix} \alpha & 1 & \mathbf{O}_{1\times n-2} & -\alpha & 0 & \mathbf{O}_{1\times(n-2)(p-1)} \end{bmatrix}$$

so that it picks up the argument of the expectations in (26) within the vector $\widetilde{Z}_t$. Also, define a $(1 \times np)$ vector ζ as

$$\underbrace{\zeta}_{1\times pn} = \begin{bmatrix} \frac{\alpha\lambda}{1-\lambda} + \alpha & 0 & \mathbf{O}_{1\times(n-2)} & -\alpha & 0 & \mathbf{O}_{1\times(n-2)(p-1)} \end{bmatrix}$$

so that it picks the LHS of (26) 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

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

or:

$$\zeta \widetilde{Z}_t = \widetilde{\delta} \tilde{\varepsilon}_t + \sum_{i=0}^{\infty} (1-\lambda)^{i+1} \, \widetilde{\delta} D^{i+1} \tilde{\varepsilon}_{t-1-i} \tag{29}$$

I'll show that

$$\sum_{i=0}^{\infty} (1-\lambda)^{i+1} \, \widetilde{\delta} D^{i+1} \tilde{\varepsilon}_{t-1-i} \approx (1-\lambda) \, \overline{x} \cdot \zeta \widetilde{Z}_{t-1} \tag{30}$$

In equation (30) the terms of summation are all 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

$$\begin{aligned} tr\left(\sum_{i=0}^{\infty} \gamma^{i+1} \widetilde{\delta} D^{i+1} \tilde{\varepsilon}_{t-1-i} \right) &= \gamma \sum_{i=0}^{\infty} \gamma^i tr\left(D^{i+1} \tilde{\varepsilon}_{t-1-i} \widetilde{\delta} \right) \\ &= \gamma \sum_{i=0}^{\infty} \gamma^i x_i tr\left(D^i \tilde{\varepsilon}_{t-1-i} \widetilde{\delta} \right) \\ &\simeq \gamma \overline{x} \cdot tr\left(\sum_{i=0}^{\infty} \gamma^i \delta D^i \tilde{\varepsilon}_{t-1-i} \right) \end{aligned} \tag{31}$$

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

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

with actual US data. 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 (31) in (29). Plus, from (28) lagged we have

$$tr\left(\sum_{i=0}^{\infty}\gamma^{i}\delta D^{i}\tilde{\varepsilon}_{t-1-i}\right)=tr\left(\zeta\widetilde{Z}_{t-1}\right) \tag{32}$$

So, plugging (32) into (31), and the result in (29) 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 VAR (27) estimates are consistent then λ_{T}^{Gmm} is also consistent.