Working papers series

WP ECON 10.09

Sticky Information and Inflation Persistence: Evidence from U.S. Data

Benedetto Molinari (U. Pablo de Olavide)

JEL Classification numbers: E31, C51

Keywords: Sticky Information, Inflation Persistence, two-stage GMM estimator

Department of Economics

Sticky Information and Inflation Persistence: Evidence from U.S. Data*

Benedetto Molinari†

October 2010

Abstract

This paper provides a novel single equation estimator of the Sticky Information Phillips Curve (SIPC), which permits to estimate the exact model without any approximation or truncation. In detail, information stickiness is estimated by employing a GMM estimator that matches the theoretical with the actual covariances between current inflation and the lagged exogenous shocks that affect firms' pricing decisions, which are considered the moments that measure inflation persistence. The main result of the paper is to show that the SIPC model can match inflation persistence only at the cost of mispredicting the variance of inflation, which is a novel finding in the empirical literature on the SIPC.

JEL Classification: E31, C51

Keywords: Sticky Information, Inflation Persistence, two-stage GMM estimator

*This paper was part of my Ph.D. thesis at UPF. I'm grateful to Albert Marcet for his guidance during that period and for valuable talks on this paper. I would like also to thank Klaus Adam, Luca Gambetti, Pau Rabanal, Jaume Ventura, Michel Reiter, Wouter den Haan, the participants at UPF macroeconomic seminars, Asset 2007 Conference, and the 2008 BoF/CEPR Conference "Expectations and the Business Cycle" for useful comments. Any remaining error is my responsibility. Financial support from Ente Luigi Einaudi (EIEF) is gratefully acknowledged.

†Department of Economics, Universidad Pablo Olavide (Sevilla, Spain – EU). Contact email: benedetto.molinari@upf.edu

1 Introduction

This paper estimates the extent of sticky information in the U.S. postwar economy. By sticky information, we refer to frictions in information flows that firms absorb to set their prices optimally. The model used to estimate information stickiness is the Sticky Information Phillips Curve [henceforth, SIPC], which was originally proposed by Mankiw and Reis (2002) [henceforth, MR] as the structural theory of inflation. The main objective of the paper is to assess the relevance of sticky information as a determinant of the actual inflation persistence observed in macroeconomic data.

In MR's view, sticky information is the key determinant of the high persistence observed in actual inflation data. Such persistence is in fact a puzzling piece of empirical evidence that conflicts with the predictions of several textbook theories of prices, such as the neoclassical model of monopolistic competitive firms with no nominal rigidities or the New Keynesian Phillips Curve model.[1] In general, if firms want to maximize profits and have rational expectations, they react to any exogenous shock by adjusting their prices as soon as they acquire information about the incoming shock, and in each period, they acquire all of the information available in the economy to identify new shocks. Hence, the effect of shocks on price changes lasts for few periods, and in a model of the aggregate supply, the only source of persistence in inflation dynamics can be the exogenous shocks (e.g. cost push shocks, monetary policy shocks, demand shocks). To overcome the lack of intrinsic persistence in inflation dynamics, MR conjectured that firms only sporadically absorb the information needed to price their goods optimally. In periods when information inflows are limited or absent, firms rely on outdated price plans. When a shock occurs only a fraction of firms adjust their prices contemporaneously, while the others delay their optimal adjustment until they become aware of the "new" shock. Thus, the overall effect of shocks on price changes lasts several periods, and inflation turns out to be a persistent process as the real data suggest.

In the SIPC model, the frequency of information updating by firms is the key parameter that controls for intrinsic inflation persistence. For a given persistence of exogenous shocks, high (low) values of the frequency of information updating − parameter λ − imply low (high) persistence of fitted inflation generated by the model. Using model simulations, Reis (2006) showed that $\lambda = 0.25$ is the best parameter for the SIPC to match the persistence of U.S. postwar inflation once a process for exogenous shocks that resembles the one observed in the actual economy is assumed. The subsequent empirical literature on the SIPC model, however, was not unanimous in its support of this calibration. Khan and Zhu (2006), Kiley (2006), Coibion (2010), and Korenok (2008) estimated $\lambda_T \in [0.15, 0.4]$ using single equation estimators, while Laforte (2006), and Mankiw and Reis (2007) estimated $\lambda_T \in [0.7, 0.85]$ using fully fledged DSGE model estimations, where the SIPC is used to model the aggregate supply curve. In this paper, such a discrepancy between the two ranges of estimates is

[1] MR presented the SIPC as an alternative theory to the New Keynesian Phillips Curve (NKPC), which was criticized because of its lack of persistence. The criticisms pointed out that (i) actual inflation responds gradually to monetary policy shocks, while NKPC implies an immediate adjustment; (ii) output losses typically accompany a reduction in inflation, while this is not true with NKPC; (iii) NKPC implies that announced disinflation causes a boom, while in a real economy the opposite is true. See Mankiw and Reis (2002).

relevant because in the second case, the SIPC model predicts little intrinsic persistence in inflation dynamics, a prediction that is at odds with the data and with the original purpose of this model.[2] This point is analyzed in detail in section 2.2.

This paper suggests that the estimates of λ vary significantly depending on which moment of the inflation process is used to estimate the SIPC model, which in facts explains the reason for the discrepancy in the estimates of λ found in the literature. In detail, I show that it is possible to generate estimates close any of the two ranges mentioned above using the SIPC model to match either persistence or conditional variance of inflation and output. I find $\lambda_T \in [0.31, 0.58]$ when using the SIPC to match the covariances between current inflation and lagged exogenous shocks, i.e., the moments that measure inflation persistence. In this case, firms are predicted to update information every 6 to 9 months, in line with the results of those papers that used single equation estimators. In contrast, I find $\lambda_T \in [0.71, 0.91]$ when using the SIPC to match the covariances between current inflation and current exogenous shocks, i.e., the conditional variance of inflation and output. In this case, firms are predicted to update their information much more frequently, about every 4 months, in line with the results obtained by Mankiw and Reis (2007) and Laforte (2006). In Section 2.2 it is shown that the SIPC model calibrated with $\lambda \simeq 0.80$ generates a fitted inflation persistence that accounts for around 2/3 to 4/5 of actual inflation persistence, while the model calibrated with $\lambda \simeq 0.25$ generates a fitted inflation variance that accounts for around 1/2 to 2/3 of actual inflation variance. Hence, I conclude that the SIPC model can explain inflation persistence only at the cost of mispredicting inflation variance.

Previous findings are obtained by employing a novel single equation estimator of the SIPC model, which is provided in Section 3.1. This estimator allows the researcher to choose explicitly which moments from the data to use to estimate the sticky information parameter. In this paper, I use as moments the covariance between current inflation and current and lagged exogenous shocks. These shocks determine firms' pricing decisions and, consequently, the resulting covariances seem to be the appropriate moments at which to estimate the SIPC parameters: the more firms are inattentive to new shocks, the longer a shock today will affect future price settings and, thus, the longer it will be correlated with price changes (i.e. inflation).

The theoretical covariances from the SIPC model are derived by writing the expectation terms that appear in the SIPC as functions of forecast errors, and then the forecast errors as functions of exogenous shocks, in a way similar to that described in Mankiw and Reis (2007) and explained in detail in Wang and Wen (2006).[3] Once the model is transformed in this way, it is easy to derive a set of orthogonality conditions based on the covariances between inflation and exogenous shocks. These orthogonality conditions are finally estimated using a two-stage estimator: first, a vector auto-regression [henceforth, VAR] model is fitted to the macroeconomic data to identify the exogenous shocks and to obtain the estimated covariances. Second, generated regressors from the VAR model are plugged into the orthogonality conditions derived from the SIPC model, which are eventually estimated with the Generalized Method of Moments to obtain estimates of the SIPC parameters. This econometric

[2]When $\lambda = 1$ the SIPC model encompasses the rational expectations model with monopolistic competition and flexible prices, which has been repeatedly shown in the literature to imply the counterfactual prediction of little intrinsic persistence in inflation dynamics.

[3]The technique used in this paper to manipulate the model is based on Molinari (2007).

strategy has two advantages with respect to other single equation estimators previously employed in the literature: (i) because the orthogonality conditions used in the estimation have a closed form solution, the infinite dimensions problem usually associated with estimations of the SIPC is avoided without any truncation or approximation of the original model; (ii) the estimation procedure allows the researcher to explicitly choose moments from the data considered appropriate to estimate the model parameters.

The rest of the paper is organized as follows: Section 2 presents the SIPC model and investigates the relationship between sticky information and inflation persistence. Section 3 presents the econometric strategy, develops the necessary steps to derive the two-stage estimator, and reports the estimation results. Section 4 performs several robustness analyses to test the results obtained in Section 3.3, and Section 5 analyzes whether the degree of information stickiness varies in the postwar sample. Finally, Section 6 concludes.

2 The Model

2.1 The Sticky Information Phillips Curve

The sticky information theory was originally developed by MR by combining elements of early papers on limited information theory by Lucas (1973) and Fischer (1977). MR analyzed a monopolistically competitive market populated by profit maximizing firms where firms are rational in the sense that they maximize profits by pricing their goods based on all of the information available to them but they receive new information only sporadically. Specifically, in every period, each firm is assumed to have a strictly positive probability $\lambda \leq 1$ of obtaining new information, e.g. about demand shocks in the market or changes in nominal marginal cost. When a firm receives new information, it sets its current price as in the standard model of profit maximizing firms with rational expectations. Otherwise, it sets the price that maximizes its profits conditional on the outdated information it possesses from previous periods. As a result, a fraction λ of firms in the market maximizes profits conditional on newly updated information, while the remaining $(1-\lambda)$ of firms sets prices according to their old price plans. In particular, MR showed that inflation π_t in this model economy evolves according to

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

where $\Delta y_t = y_t - y_{t-1}$ is the growth rate of the output gap, α measures the steepness of the aggregate demand curve, and λ is the probability that the agent's information is updated in period t. They called equation (1) the Sticky Information Phillips Curve.[4]

In the SIPC, inflation is persistent because current inflation depends on past periods expectations, and higher weights $(1-\lambda)^j$ on past expectations, i.e. lower λ, results in more persistent inflation. Intuitively, the mechanism that generates persistence is as follows: suppose that in period t, a shock ε_t increases the output gap, and the knowledge about this

[4]See Mankiw and Reis (2002).

shock is included in period t information, $\varepsilon_t \subset \Omega_t$. Then, inflation increases contemporaneously because of the trade-off term $\frac{\alpha\lambda}{1-\lambda}y_t$. In the next period $t+1$, a fraction λ of agents becomes aware of ε_t that they did not observe in t and adjusts their prices accordingly. According to (1), inflation π_{t+1} increases because of the term $E\left[\pi_{t+1} + \alpha\Delta y_{t+1} \mid \Omega_{(t+1)-1-j}\right]$, which is positive for $j = 0$, i.e., for those firms that in $t+1$ obtained the information dating to t. The same happens in $t+2$, when a fraction $\lambda(1-\lambda)$ of firms becomes aware of the shock, and so forth in all of the following periods $t+j \; for \; j > 1$. As apparent from equation (1), the effect of past shocks on current inflation fades out at the rate $(1-\lambda)^j$. Hence, in this model a shock today affects all future inflation levels, thus implying that the inflation process is serially correlated for many periods as the real data suggest.

2.2 Sticky Information and Inflation Persistence

In the SIPC model, the key parameter that controls for endogenous (intrinsic) persistence in fitted inflation is the frequency of information updating by firms, λ. This point can be shown simulating the SIPC model repeatedly for all the values of $\lambda \in (0, 1]$.[5] Figure 1 plots the ratio of the first order autocovariance function [henceforth, $acf(1)$] of fitted inflation to the one of actual inflation. From Figure 1 it is apparent that high (low) values of λ imply low (high) persistence of the fitted inflation. Notice that the overall degree of inflation depends not only on the value of λ but also on the persistence of the exogenous shock process. In particular, in the limiting case of $\lambda = 1$, all of the persistence of fitted inflation exhibited in the SIPC model comes from exogenous shocks, which in our example roughly coincides with the persistence of the $AR(1)$ process.

MR assumed that producers receive new information in every period with an exogenous probability λ equal to 0.25, thus setting firms' average information duration as one year.[6] Using this calibration, MR showed that fitted inflation from the SIPC model responded gradually to several exogenous shocks, e.g., demand shocks and monetary policy shocks. Reis (2006) proposed a validation test of MR's calibration based on model simulations. He showed that simulated inflation from the SIPC model calibrated with $\lambda = 0.25$ does a good job in matching some selected moments of the actual aggregate distribution of prices including the first autocovariance function of inflation, which was used as measure of inflation persistence.

Using previous result, Reis concluded that the SIPC model can generate dynamics of inflation as persistent as the one observed in actual data. Yet, the process of exogenous shocks used by Reis to simulate the SIPC was calibrated ad-hoc in order to match some long-run facts of the U.S. postwar economy, and not estimated jointly with the SIPC model. In this fashion, proper econometric estimations of the SIPC model based on macroeconomic data has been provided in several papers, e.g., Khan and Zhu (2006), Kiley (2006) Korenok (2008), Coibion (2010), Laforte (2007), Mankiw and Reis (2007). Among these authors, however, there was no consensus on the estimate of λ. Khan and Zhu (2006), Kiley (2006)

[5]Following Reis (2006), I use a simple univariate $AR(1)$ process for the exogenous shocks, and I take the first order autocovariance function as a measure of inflation persistence.

[6]The average duration of information is $D = (1-(1-\lambda))^{-1}$. Thus, in the quarterly model used by MR, $\lambda = 0.25$ implies $D = 4$, i.e., four quarters.

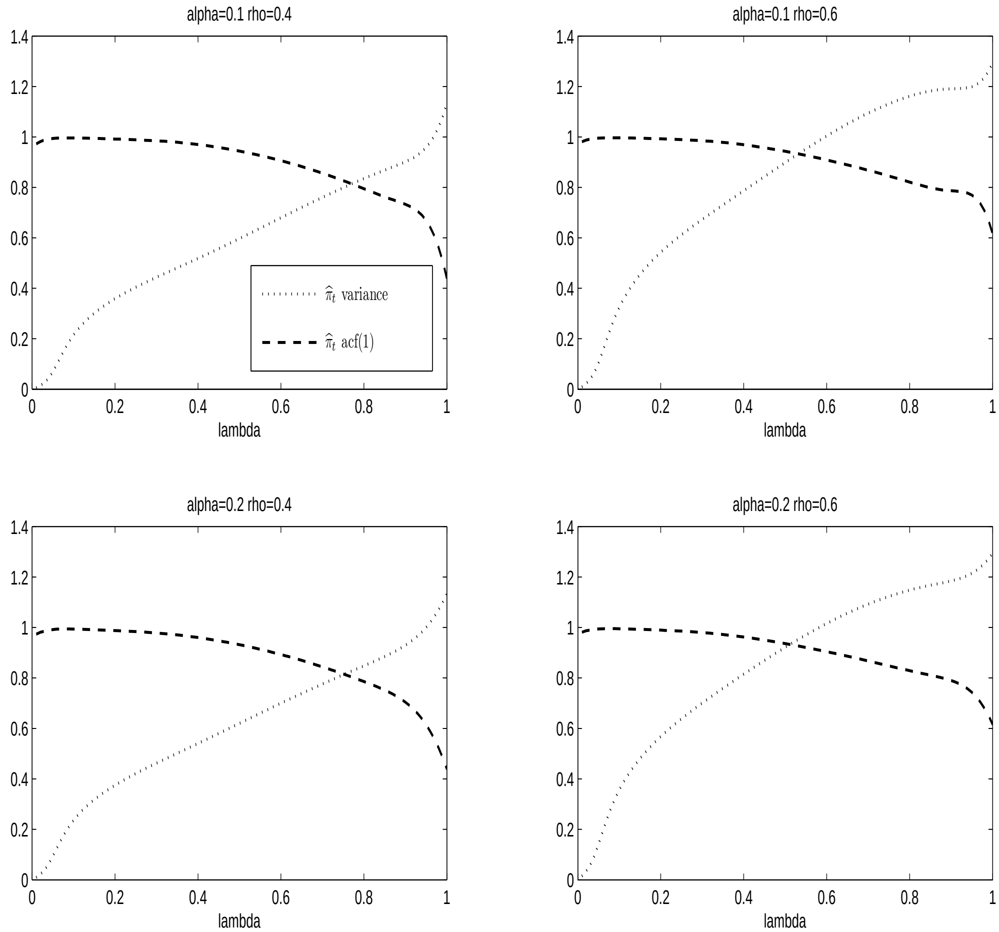

Figure 1: **Simulated moments of fitted inflation from the SIPC model.** Fitted Inflation $\widehat{\pi}_t$ is generated from SIPC model varying the sticky information parameter $\lambda \in (0, 1]$. *Alpha* is defined as in equation (1), *rho* is the autoregressive parameter in the AR(1) process of exogenous shocks. Dashed line represents the first-order autocorrelation function of $\widehat{\pi}_t$, *acf(1)*, and dotted line is the variance of $\widehat{\pi}_t$.

Korenok (2008) and Coibion (2010) estimated $\lambda_T \in [0.15, 0.4]$ using single equation estimators, while Mankiw and Reis (2007) and Laforte (2007) estimated $\lambda_T \in [0.7, 0.85]$ using multiple equation estimations.[7]

The results obtained by the first set of papers is intuitively clear once noting that single equation estimators match persistence in the data by construction. These papers estimated the SIPC (1) by truncating the infinite sum of expectations at $t - j_{\max}$ and then substituting the remaining expectations terms with the predictions of a VAR model set ad-hoc to forecast

[7]These are estimations of structural DSGE models where the SIPC model is used to model the aggregate supply.

inflation and output gap. So, their estimated specification was

$$\pi_t = \frac{\alpha\lambda}{1-\lambda} y_t + \lambda proj_{t-1}\left(\pi_t + \alpha\Delta y_t\right) + ... + \lambda(1-\lambda)^{j_{\max}} proj_{t-1-j_{\max}}\left(\pi_t + \alpha\Delta y_t\right)$$

Now, because the $proj_{t-j}\left(\pi_t + \alpha\Delta y_t\right)$ is a linear combination of lagged inflation and output gap (plus past values of other variables possibly included in the VAR), they basically regressed current inflation on $t - j_{\max}$ lags of inflation and output gap, i.e. the covariances between inflation and lagged inflation that can be used as measure of inflation persistence. Hence, it seems reasonable that they found evidence in support of Reis' (2006) calibration.

What seems puzzling, then, is the second set of estimates, i.e., $\lambda_T \in [0.7, 0.85]$. If the SIPC model were the true data generating process (DGP), we should find similar estimates of λ when matching any moment from the data. It is not immediate to understand the reason of such discrepancy in terms of moments matching estimator because Mankiw and Reis (2007) and Laforte (2007) estimated fully fledge DSGE model with Maximum Likelihood estimators. In this paper, however, I show that it is possible to replicate their results using the SIPC to match the covariances between current inflation and current exogenous shocks, i.e., the variance of inflation conditional to shocks to prices and output. Figure 1 shows that the SIPC model calibrated with $\lambda \simeq 0.80$ generates a fitted inflation persistence that accounts for around 2/3 to 4/5 of actual inflation persistence.[8] Moreover, the SIPC model calibrated to match actual inflation persistence, $\lambda \simeq 0.25$, generates a fitted inflation variance that accounts for around 1/2 and 2/3 of the actual inflation variance. Such pattern in the dynamics of inflation generated by the SIPC model could explain why, when λ is estimated using the SIPC to match the conditional variance of inflation, the resulting estimates are statistically different from those of Reis (2006), Khan and Zhu (2006), Kiley (2006) and Korenok (2008).

3 The Estimation

3.1 The econometric strategy

To estimate the parameters of the SIPC model, I assume that the dynamics of inflation and the output gap result from the interaction of n macroeconomic variables, which are defined as the elements of a covariance-stationary vector process Z_t. This assumption imposes very little structure on the processes of inflation and output gap, and it permits to derive a useful result from the SIPC model, which is provided in the following proposition.

Proposition 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 the SIPC (1) implies:*

$$\frac{\alpha\lambda}{1-\lambda} y_t + \alpha\Delta y_t = \sum_{i=0}^{\infty} (1-\lambda)^i \, \delta A_i \varepsilon_{t-i} \tag{2}$$

[8] This result is clearly restricted to the case of the exogenous shock process assumed during the SIPC simulation, and therefore, provides only a first approximation explanation of the results presented in next section.

where A_i and ε_t are, respectively, the dynamic multipliers and the uncorrelated errors of the MA representation of Z_t. δ is a $(1 \times n)$ row vector that defines the process $(\pi_t + \alpha \Delta y_t)$ using the elements of Z_t.

Proof. See Appendix A. □

Multiplying (2) by a vector of lagged errors ε_{t-i} and taking the expectations, I obtain

$$\begin{aligned} E\left[\left(\frac{\alpha\lambda}{1-\lambda} y_t + \alpha \Delta y_t\right)(\delta \varepsilon_{t-i})'\right] &= (1-\lambda)^i \, \delta A_i \Sigma \delta' \\ for \quad i &= 0, \ldots, L \end{aligned} \tag{3}$$

where $\Sigma \equiv E\left[\varepsilon_t \cdot \varepsilon_t'\right]$ is the Variance-Covariance (VCV) matrix of the errors ε_t.[9]

Equations (3) constitute a system of orthogonality conditions [henceforth, o.c.] that can be used to estimate the parameters of the SIPC model using the Generalized Method of Moments [henceforth, GMM]. In detail, each equation in (3) defines the theoretical covariance between Z_t and ε_{t-i} in the SIPC model, and the vector of the RHS terms in (3) can be interpreted as a linear combination of the vector of the first i lags of the Impulse Response Functions of π_t and Δy_t w.r.t. ε_t.[10] Intuitively, in the SIPC model the Impulse Response Functions of inflation depend on the shocks affecting the output gap and inflation, which are the driving force of prices, and on λ, which measures how many firms are attentive to the shocks, i.e. how rapidly the effect of the shocks on prices fades out.

The estimation of (3) is complicated by the fact that $\{\varepsilon_t, A_i, \Sigma\}$ are unknown regressors. I overcome this problem by employing a two-stage estimator that works as follows. First, it fits a VAR model to Z_t to obtain consistent estimates of $\{\varepsilon_t, A_i, \Sigma\}$. Second, it estimates the o.c. (3) with the GMM using $\left\{\widehat{\varepsilon}_t(\beta), \widehat{A}_i(\beta), \Sigma_T(\beta)\right\}|_{\beta=\beta_T^{VAR}}$ as regressors.[11] This econometric strategy implies that some data used in the second stage estimation are regressors generated from the first stage estimation. Consequently, the asymptotic standard errors of λ_T^{2s} calculated from the GMM algorithm needs to be adjusted to account for the variance of the stochastic regressors. I do this by deriving the adjusted variance analytically rather than following the more common Monte Carlo or bootstrapping approaches. The analytical approach, which is not feasible in many cases, can be easily implemented here because the stochastic regressors are functions of the VAR(p) parameters, whose variance is known and estimated in the first stage estimation. The value of the adjusted variance of the two-stage estimator, λ_T^{2s}, is given in the following proposition.

[9] Equation (3) follows after multiplying (2) by $(\delta\varepsilon_{t-i})'$ and taking the expectation conditional on information at time t. It uses the fact that the errors are uncorrelated, i.e. $E\left[\varepsilon_t \varepsilon_{t-i}\right] = 0$ for $i \geq 1$. Later on in the text the orthogonality conditions (3) may be referred to as $g_{1,t}$. See Appendix B.

[10] This follows immediately from the assumption that Z_t is ergodic (see Proposition 1).

[11] If $\left\{\widehat{\varepsilon}_t, \widehat{A}_i, \Sigma_T\right\}$ are consistent estimates of $\{\varepsilon_t, A_i, \Sigma\}$, it can be shown that the sample analog:

$$\frac{1}{T}\sum_{t=1}^{T}\left[\left(\frac{\alpha\lambda}{1-\lambda} y_t + \alpha \Delta y_t\right)(\delta \widehat{\varepsilon}_{t-i})' - (1-\lambda)^i \, \delta \widehat{A}_i \Sigma_T \delta'\right] = 0$$

converges almost surely to the population moment (3).

Proposition 2. *Let $g_{1,t}$ be the vector of orthogonality conditions defined in (3), $g_{2,t}$ be the vector of orthogonality conditions used to estimate the VAR(p) model in the first stage, and Σ_{g_h} be the VCV matrix of the o.c. $g_{h,t}$ for $h = \{1, 2\}$. Also, denote $\beta = vec\left(\left[B_1', \ldots, B_p'\right]'\right)$ where B_j are the matrices of parameters in the VAR(p) model, and $vec(\cdot)$ the column stacking operator. Then, if $E\left(g_{1,t} \cdot g_{2,t}'\right) = 0$, the variance of λ_T^{2s} is:*

$$V(\lambda_T^{2s}) = \Bigg[\left(TV_{na}(\lambda_T^{2s})\right)^{-1} - E\frac{\partial g_{1,t}'}{\partial\lambda}\Sigma_{g_1}^{-1}E\frac{\partial g_{1,t}}{\partial\beta'} \cdot \left(E\frac{\partial g_{1,t}'}{\partial\beta}\Sigma_{g_1}^{-1}E\frac{\partial g_{1,t}}{\partial\beta'} + E\frac{\partial g_{2,t}'}{\partial\beta}\Sigma_{g_2}^{-1}E\frac{\partial g_{2,t}}{\partial\beta'}\right)^{-1} \cdot E\frac{\partial g_{1,t}'}{\partial\beta}\Sigma_{g_1}^{-1}E\frac{\partial g_{1,t}}{\partial\lambda}\Bigg]^{-1}/T \quad (4)$$

where $\frac{\partial g_{1,t}}{\partial\lambda}$ and $\Sigma_{g_1}^{-1}$ are, respectively, the jacobian and the weighting matrix of the second stage GMM estimator, and $\left(E\frac{\partial g_{2,t}'}{\partial\beta}\Sigma_{g_2}^{-1}E\frac{\partial g_{2,t}}{\partial\beta'}\right)$ is the VCV matrix of the VAR(p) coefficients. Finally, $\frac{\partial g_{1,t}}{\partial\beta'}$ is the vector of derivatives of (3) with respect to β, and $V_{na}(\lambda_T^{2s})$ is the (not-adjusted) variance of the second stage estimator.

Proof. See Appendix B. □

In equation (4) the adjusted variance of λ_T^{2s} is written as function of the unadjusted variance, $V_{na}(\lambda_T^{2s})$ and, as expected, we find that $V\left(\lambda_T^{2s}\right) \geq V_{na}\left(\lambda_T^{2s}\right)$.[12] It is useful to notice that the only additional task required to compute (4) is the derivation of $\frac{\partial g_{1,t}}{\partial\beta'}$ and its evaluation at $\{\beta_T^{VAR}, \lambda_T^{2s}\}$ because $\frac{\partial g_{1,t}}{\partial\lambda}$ and $\Sigma_{g_1}^{-1}$ are delivered by the GMM algorithm that computes λ_T^{2s}, and $\left(E\frac{\partial g_{2,t}'}{\partial\beta}\Sigma_{g_2}^{-1}E\frac{\partial g_{2,t}}{\partial\beta'}\right)$ is generated by the algorithm that estimates the VAR(p) model.

The adjusted variance (4) uses the assumption that $E\left(g_{1,t} \cdot g_{2,t}'\right) = 0$. In this case the two stages estimator implies no loss of efficiency relative to a one-stage GMM estimator that jointly estimates $[\lambda, \beta']'$ pooling together $g_{1,t}$ and $g_{2,t}$.[13] This is intuitively clear because λ does not appear in the reduced form equations of the VAR(p) model, and if the residuals of the first stage estimation are not correlated with the residuals of the second stage estimation, then there is no extra information in the VAR(p) model that can be exploited to pin down λ. In Section 4.2 the assumption of zero covariance between residuals will be relaxed to investigate its effect on the estimates of λ.

3.2 First stage: VAR(p) estimation of Z_t

This section provides estimates of the VAR(p) model used to generate the regressors for the second stage estimation. From the perspective of the SIPC model, the VAR(p) model comprises the forecasting technology used by firms to make predictions about future output

[12]This follows immediately from the fact that the quadratic terms in (4) are all positive definite.
[13]See Appendix B.

gap and inflation. I estimate two specifications of this VAR(p) model: (i) the baseline specification where the vector Z_t includes inflation, the output gap and the interest rate; and (ii) a more extended specification where the vector Z_t includes the variables most relevant to forecast inflation and the output gap according to Stock and Watson (2003).[14] I will refer to this second specification as the *min RMSE*.

In details, I estimate the reduced form coefficients B_j of:

$$Z_t = \sum_{j=1}^{p} B_j \cdot Z_{t-j} + \varepsilon_t \tag{5}$$

where

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

and X_t can be either $X_t = i_t$, or a $(n - 2 \times 1)$ vector that includes: short term interest rate (Fed Fund Rate); the term spread (10 years Government bond minus short term interest rate); the real Stocks Price Index (S&P500, deflated by CPI); the IMF price index of commodities; real money (real M2 minus small time deposits); unemployment rate; total capacity utilization rate (TCU).

The data sample includes the period from 1957q1 to 2005q4 (196 observations).[15] All of the variables are taken in logs except for unemployment, TCU, and interest rates. The variables have been detrended or first differences have been taken when necessary,[16] and, accordingly, the series used in the VAR(p) are all stationary. Finally, the order p is chosen to be the minimum order of lags for the VAR(p) residuals to be not serially correlated. These conditions together assure that the VAR(p) estimates are consistent. The two specifications − *baseline* and *minRMSE* − are estimated twice, with inflation first measured as the CPI and then as the implicit GDP deflator. Eventually, the triplet $\left\{\widehat{\varepsilon}_t, \widehat{A}_i, \Sigma_T\right\}$ is computed and collected for the four models and stored for the second stage estimation.

3.3 Second stage: the GMM estimation

This section present the GMM estimation of the system of o.c. (3). To control for the small sample bias problem that typically affect nonlinear GMM estimates, I use two alternative specifications of the orthogonality conditions: the first is (3) multiplied by $(1 - \lambda)$, and the second is (3) multiplied by $\left(\frac{1-\lambda}{\alpha\lambda}\right)$. These are referred to as (a) and (b) in subsequent tables. The data sample includes the period from 1958q4 to 2005q4 (189 observations).[17]

In the estimation, the parameter α is fixed and λ is estimated alone. In the original SIPC model, α is a reduced-form parameter that captures the degree of strategic complementarity among firms' prices in the aggregate supply,[18] and is function of the intertemporal

[14]Stock and Watson (2003) analyzed the contribution of several variables to forecast inflation and output gap.

[15]All data used in this paper come from the FRED-II database issued by the Federal Reserve Bank of St. Louis.

[16]The output gap is the real GDP detrended with the HP filter.

[17]The data sample used in the second stage estimation is shorter than the one used in the first stage because 7 observations are lost when obtaining the VAR(6) estimates.

[18]See Reis (2006).

elasticity of substitution, Frisch elasticity of the labor supply, and the elasticity of the demand of single-variety goods in the monopolistic competition setup. Now, because no data on consumption, labor, or markup is used in this estimation, parameter α would probably behave in the GMM algorithm as a free parameter, eventually adjusting in such a way the algorithm improves the matching of the moments, rather than itself matching the actual degree of strategic complementarity in the aggregate supply. For this reason, I estimate the o.c. (3) by fixing α, and then I check whether the results are sensitive to such restriction. According to the literature,[19] a standard calibration for α lies in the interval $\alpha \in [0.1, 0.2]$. Thus, I fix $\alpha = 0.2$ and, in the next section, I re-estimate the model using $\alpha = 0.1$.

Table 1: **Estimates of the SIPC matching the first L covariances**

O.C. (3)	$i = 0, ..., L$	J-statistic and (p-val)			
Restricted	α=.2	L=2	L=4	L=6	L=12
GDP defl. as π	(a)	19.68	23.21	23.64	79.23
VAR specif.		(0.00)	(0.00)	(0.00)	(0.00)
$\{\Delta y_t, \pi_t, i_t\}$	(b)	11.41	12.15	12.71	23.41
		(0.00)	(0.01)	(0.04)	(0.02)
GDP defl. as π	(a)	14.03	14.20	16.09	54.78
VAR specif.		(0.00)	(0.00)	(0.01)	(0.00)
minRMSE	(b)	9.62	11.27	12.59	22.06
		(0.00)	(0.02)	(0.04)	(0.03)
CPI as π	(a)	27.63	31.31	31.97	47.39
VAR specif.		(0.00)	(0.00)	(0.00)	(0.00)
$\{\Delta y_t, \pi_t, i_t\}$	(b)	11.59	11.68	11.72	14.93
		(0.00)	(0.01)	(0.06)	(0.24)
CPI as π	(a)	9.40	10.22	11.00	34.75
VAR specif.		(0.00)	(0.03)	(0.08)	(0.00)
minRMSE	(b)	6.15	6.47	6.81	18.11
		(0.04)	(0.16)	(0.33)	(0.11)

2-stage GMM with optimal weighting matrix. U.S. data, sample 1958q4 – 2005q4. Output gap filtered with HP filter. *adj. s.e.* are Newey-West standard errors adjusted for stochastic regressors. p-values in parenthesis. J-statistic is Hansen's test of over-identifying restrictions (with L degree of freedom [d.o.f.]).

The GMM estimation of o.c. (3) for $i = 0, ..., L$ delivers one main result: the rejection of the null hypothesis in Hansen's test of over-identifying restrictions. As we can see in Table 1, this result is broadly confirmed for all the specifications tested, i.e., when changing the order L up to which lagged covariances are used – either 2, 4, 6, or 12 – or the inflation index employed – CPI or GDP deflator –, or the specification of the VAR(p) model used to generate the regressors – baseline or *min RMSE*. This result indicates that the data seem to reject the estimated theoretical moment. Usually this is taken as evidence against

[19]See Reis (2006) and Woodford (2002), ch. 2.

the selected set of instruments. In present case, because the estimated moments are the covariances between inflation and lagged shocks computed using actual data and, therefore, describe the actual inflation process, the rejection of Hansen's test can be taken as evidence for the inability of the SIPC model to explain actual inflation dynamics.

In particular, previous result shows that no value of λ exists such that the theoretical moments from the SIPC jointly match the actual moments from the data. This finding is novel in the literature, and I suggest that this may be because the SIPC model can explain actual persistence only at the cost of mis-predicting actual variance. For this reason, previous attempts to estimate the o.c. (3) failed. In the rest of this section, I provide evidence to support this claim showing that the SIPC model matches the o.c. regarding inflation persistence for a value of λ different than the one needed to match the o.c. that regard inflation variance. To illustrate this point, I first estimate λ using the o.c. (3) with $i = 1, ..., L$, thus discarding the moment that reflects the contemporaneous covariance. The econometric strategy used is the same as that employed above. Table (2) reports the results.

Table 2: **Estimates of λ using lagged covariance moments**

O.C. (3)	$i = 1, ..., 6$	λ_T^{2s}	MR calib.	t-stat	RE calib.	t-stat	J-stat
Restricted	$\alpha = .2$	(adj. s.e.)	$H_0 : \lambda$	(p-val)	$H_0 : \lambda$	(p-val)	(p-val)
	(a)	0.35	0.25	1.62	1	-9.83	2.22
GDP deflator		(0.065)		(0.10)		(0.00)	(0.81)
baseline	(b)	0.36	0.25	1.66	1	-8.74	1.95
		(0.072)		(0.10)		(0.00)	(0.85)
	(a)	0.38	0.25	1.47	1	-6.61	2.72
GDP deflator		(0.092)		(0.14)		(0.00)	(0.74)
minRMSE	(b)	0.40	0.25	1.48	1	-5.72	2.37
		(0.104)		(0.13)		(0.00)	(0.79)
	(a)	0.47	0.25	3.74	1	-8.60	2.70
CPI		(0.060)		(0.00)		(0.00)	(0.74)
baseline	(b)	0.49	0.25	3.87	1	-8.14	2.24
		(0.062)		(0.00)		(0.00)	(0.81)
	(a)	0.54	0.25	2.96	1	-4.60	3.38
CPI		(0.099)		(0.00)		(0.00)	(0.64)
minRMSE	(b)	0.57	0.25	3.15	1	-4.05	2.74
		(0.104)		(0.00)		(0.00)	(0.73)

Estimates of λ_T^{2s} are obtained from 2-stages GMM with optimal weighting matrix and o.c. as indicated in row 1. U.S. data, sample 1958q4 – 2005q4. Output gap filtered with HP filter. *adj. s.e.* are Newey-West HAC standard errors adjusted for stochastic regressors. p-values in parenthesis. J-stat is Hansen's test of overidentifying restrictions (with 5 d.o.f.).

The theoretical moments from the SIPC model now describe inflation persistence well, and we can never reject the null hypothesis of over-identifying restrictions. In all the specifications, the estimates of the frequency of information updating lie in the range assumed by the theory, i.e., within the $(0, 1]$ interval. More precisely, λ_T^{2s} lies in the range $[0.35, 0.57]$, implying an average duration of around 6 to 9 months for the information. In column 4 of Table 2, I test MR's calibration, reporting the t-statistic and the corresponding p-value for the null hypothesis of $\lambda = 0.25$. The null is accepted in most of the cases. Intuitively, the meaning of this result is clear. Reis (2006) showed that, for a reasonable process of the exogenous shocks, $\lambda = 0.25$ is the correct value for the SIPC model to reproduce the persistence of actual U.S. postwar inflation. Therefore, if we force the model to match the moments from the data that measure inflation persistence, and if the process for the shocks extrapolated from the data using the VAR (5) is similar to the process used by Reis, then the most appropriate value of λ is likely to be 0.25, as proposed by Reis. In general, this result is in line with those obtained in prior papers that estimated the SIPC model using single equation estimations. This is probably because these studies used similar information from the data to estimate λ, as shown in Section 2.2.[20]

Finally, it is worth noticing that λ_T^{2s} appears to be sensitive to the magnitude of the exogenous shocks. Larger λ_T^{2s} occurs concurrently with smaller forecast errors $\{\widehat{\varepsilon}_t\}_{t=1}^{T}$, as seen when comparing the estimates that use the *minRMSE* versus *baseline* specification.[21] This result can probably be explained by the difference between the Impulse Response Functions computed in the *baseline* and in the *minRMSE* specifications. In the *minRMSE* specification, a fraction of the total persistence of actual inflation is explained by lags of the other variables included in the VAR, e.g. spread term, SP500 index, etc. In the *baseline* specification, all of the persistence is explained by lags of π_t and y_t, which work as proxies for other variables. Hence, the estimated persistence conditional on the shocks of inflation and output gap is higher in the baseline specifications, and accordingly, we find that a lower value of λ is needed for the SIPC model to match persistence when using the baseline VAR(p) regressors in the estimation of o.c. (3).

At this point, the question that naturally follows is whether the SIPC model can match the estimated conditional variance of δZ_t alone and, if so, for which value of λ. To this end, I estimate the o.c. (3) with $i = 0$, discarding the o.c. related to lagged covariances and keeping only the contemporaneous one. To obtain more precise estimates, the single orthogonality

[20]For a more precise comparison, I estimated λ using the same information as used in previous papers, which turns out to be the information contained in the covariances between inflation and the first L lags of δZ_{t-i}. In this case, using equation (2) and lagged δZ_{t-i} as instruments, I derive and estimate the following orthogonality conditions:

$$\begin{aligned} E\left[\left(\frac{\alpha\lambda}{1-\lambda}y_t + \alpha\Delta y_t\right)(\delta Z_{t-i})'\right] &= \sum_{j=0}^{\infty}(1-\lambda)^{i+j}\,\delta A_{i+j}\Sigma A_j'\delta' \\ for \quad i &= 1, ..., L \end{aligned} \tag{6}$$

where the infinite summation on the RHS of (6) is simulated at each step of the GMM algorithm, replacing ∞ with $J_{max} = 120$. In the estimation of (6), λ_T^{2s} ranges between $[0.30, 0.41]$, a result in line with that obtained using lagged ε_t as instruments and very close to the estimates of Khan and Zhu (2006), Kiley (2006), and Korenok (2008).

[21]The *min RMSE* specification predicts Z_t better with respect to the baseline specification because it uses more information. As a result, the residuals $\widehat{\varepsilon}_t$ from the *min RMSE* specification are smaller than the ones in the *baseline*.

condition is multiplied by a vector of instruments x_t that contains all variables dated $t-1$ and before.[22] Under the standard assumption that the errors ε_t defined in Proposition 1 are i.i.d., it holds that

$$E\left[\left(\left(\frac{\alpha\lambda}{1-\lambda}y_t + \alpha\Delta y_t\right)(\delta\widehat{\varepsilon}_t)' - \delta\Sigma_T\delta'\right)\cdot x_t\right] = 0 \tag{7}$$

The system of orthogonality conditions (7) is estimated using the same procedure employed above to estimate (3). In this case the GMM point estimate of λ_T^{2s} coincides with the estimate of the non-linear IV estimator, but they have a smaller variance because the weighting matrix in the GMM is chosen optimally to be the inverse of the variance of moments. The results are reported in Table 3.

Table 3: **Estimates of λ using the contemporaneous covariance moments**

O.C. (7)	$i=0$	λ_T^{2s}	MR calib.	t-stat	RE calib.	t-stat	J-stat
Restricted	$\alpha = .2$	(adj. s.e.)	$H_0:\lambda$	(p-val)	$H_0:\lambda$	(p-val)	(p-val)
	(a)	0.75	0.25	6.72	1	-3.33	22.21
GDP deflator		(0.074)		(0.00)		(0.00)	(0.22)
baseline	(b)	0.86	0.25	7.70	1	-1.69	15.29
		(0.079)		(0.00)		(0.09)	(0.64)
	(a)	0.71	0.25	4.64	1	-2.90	22.62
GDP deflator		(0.099)		(0.00)		(0.00)	(0.20)
minRMSE	(b)	0.83	0.25	5.70	1	-1.55	15.55
		(0.103)		(0.00)		(0.12)	(0.62)
	(a)	0.85	0.25	9.88	1	-2.35	19.61
CPI		(0.061)		(0.00)		(0.01)	(0.35)
baseline	(b)	0.90	0.25	10.46	1	-1.50	11.99
		(0.062)		(0.00)		(0.13)	(0.84)
	(a)	0.85	0.25	7.37	1	-1.69	21.99
CPI		(0.082)		(0.00)		(0.09)	(0.23)
minRMSE	(b)	0.91	0.25	7.74	1	-1.03	14.65
		(0.085)		(0.00)		(0.30)	(0.68)

Table 2. Estimates of λ_T^{2s} are obtained from 2-stages GMM with optimal weighting matrix and o.c. as indicated in row 1. Data sample 1958q4 – 2005q4. HP filter for output gap. Newey-West HAC standard errors adjusted for stochastic regressors. p-values in parenthesis. J-stat is Hansen's test of overidentifying restrictions (18 d.o.f.).

The theoretical moments from the SIPC model fit relatively well to the estimated conditional variance, and we can never reject the null hypothesis of over-identifying restrictions.[23]

[22] Following similar GMM estimates in the literature, e.g. Galí and Gertler (1999), I use 19 instruments: a constant, 4 lags of inflation, 4 lags of output gap, and 2 lags of unemployment rate, interest rate, marginal cost, money growth, and term spread.

[23] In the estimations with IV estimators, such as the one presented here, standard distributions for hypoth-

The estimates of λ_T^{2s} are quite precise and range in the interval $[0.71, 0.91]$. This value implies that the average firm's information duration ranges from 3.3 to around 4 months (ca. 100 to 120 days), which is significantly shorter than the duration inferred from Table 2. Column 3 of Table 3 reports the t-statistic and the p-value for the null hypothesis that $\lambda_T^{2s} = 0.25$. In this estimation Reis's calibration is rejected for all of the specifications, suggesting that it is not an appropriate calibration of λ for the SIPC to match the conditional variance of δZ_T.

Finally, it is worth noticing that the estimates of λ_T^{2s} appear quite close to the upper bound of 1, for which the SIPC model encompasses the standard monopolistic competition model with flexible prices and rational expectations. Thus, to test the sticky information theory against this rational expectations model, in column 6, I report the t-statistic and the p-value for the null hypothesis of $\lambda = 1$. The null is rejected in most of the specifications at the 5% level, but it is accepted in more than half of them at the 1%. Overall, the evidence is mixed, and we cannot conclude anything about the rational expectations equilibrium from this estimation.

4 Robustness analysis

4.1 Empirical robustness

This section tests the robustness of previous results along several dimensions. First, I check whether the estimates of λ_T^{2s} are sensitive to changes in the fixed value of α. As mentioned in Section 3.3, parameter α has typically been calibrated in the literature within the $[0.1, 0.2]$ interval. Therefore, I re-estimate the model using $\alpha = 0.1$.[24]
Second, I test the robustness of previous results when changing the filter used to obtain the output gap series. Two-sided filters, such as the Hodrick-Prescott used in Section 3.3, may give inconsistent GMM estimates because they violate the assumption that lagged instruments must be uncorrelated with the residuals of the orthogonality conditions. To check whether this issue actually affects the estimates in the previous section, I re-estimate the model using a quadratic detrend (QD) filter instead of the Hodrick-Prescott to obtain the output gap series.

The results for the robustness exercise over the value of α are reported in Table 6, and the ones for the QD filter in Table 7. The evidence from Section 3.3 is broadly confirmed. Regardless which type of filter or which value of α is used, the SIPC model can never match all of the o.c. (3) simultaneously. As in previous estimations, however, the model does a good job of matching a subset of moments at one time, once we separate the conditional variance from lagged covariances. Again, the null hypothesis of $\lambda = 0.25$ is rejected when matching the contemporaneous covariance, and accepted in most of the specifications when matching lagged covariances.

esis testing with are reliable only if the instruments used are not weak. However, because in the literature it is still unclear how to check for weak instruments in nonlinear estimators with possibly non-spherical residuals, in what follows I rely on standard distributions for hypothesis testing and neglect this issue.

[24] I also estimated the case with $\alpha = 0.15$ to control for a possible non-monotonic effect of α on λ_T^{2s} within the $[0.1, 0.2]$ interval. These results are in line with those for $\alpha = 0.1$ presented in Table 6, and are available upon request.

4.2 Methodological robustness: One-stage GMM estimator

The inference presented in Section 3.3 is drawn under the assumption that the residuals of the two-stage estimations are mutually uncorrelated. Relaxing this assumption may affect the standard errors of λ_T^{2s} and consequently change previous conclusions. In general, if the residuals of the two stages were correlated, then we could attain a more efficient estimator of λ by jointly estimating λ and the VAR(p) parameters β instead of using the two-stage estimator. In addition, notice that the GMM estimator λ_T obtained from the joint estimation of $\{\lambda, \beta\}$ would differ from the two-stage estimator not only because of its higher efficiency if $E\left(g_{1,t} \cdot g_{2,t}'\right) = 0$, but also because the estimator β_T^{VAR} used to obtain λ_T^{2s} does not coincide with the optimal GMM estimator β_T, which estimates the VAR(p) parameters taking into account the o.c. (3) as a cross-equations restriction on the equations of the VAR(p) model. Hence, a different value of β_T can affect the estimates from λ_T, possibly changing the results obtained in Section 3.3. In order to deal with both these concerns about the two-stage estimator, in this section I derive and estimate the optimal one-stage GMM estimator $[\lambda_T,\ \beta_T']'$.

The one-stage GMM estimator entails a non linear optimization over a large set of parameters, which is computationally highly intense and tends to raise problems in the minimization algorithm of the GMM. Moreover, the dynamic multipliers A_i that appear in equation (2) have no closed solution as function of the VAR(p) coefficients and therefore should be computed numerically at each iteration of the minimization algorithm. Thus, to simplify the estimation algorithm, I assume that $Z_t = \{\Delta y_t, \pi_t\}$ and that a linear combination of demeaned inflation and output gap growth follows a univariate second order autoregressive process. Specifically,

$$S_t = \phi_1 S_{t-1} + \phi_2 S_{t-2} + \varepsilon_t \tag{8}$$

where $S_t \equiv (\overline{\pi}_t + \alpha \Delta \overline{y}_t)$, $\overline{x}$ indicates a demeaned variable x, and ε_t are *i.i.d.* errors with variance σ_ε^2. From the perspective of the SIPC model, equation (8) is equivalent to the assumption that firms use an $AR(2)$ univariate model to forecast future inflation and output gap.[25]

Using equation (8), the dynamic multipliers that appear in the o.c. (3) can be written in closed form as functions of the AR(2) parameters, which in turn allows to estimate jointly $\{\lambda, \phi_1, \phi_2, \sigma_\varepsilon^2\}$ and thus to implement the optimal estimator. In detail, the estimated system of o.c. is now

$$E\begin{bmatrix} \left(\frac{\alpha\lambda}{1-\lambda} y_t + \alpha \Delta y_t\right) \cdot \varepsilon_{t-i} - (1-\lambda)^i A_i \sigma_\varepsilon^2 \\ \begin{pmatrix} S_{t-1} \\ S_{t-2} \end{pmatrix} \cdot (S_t - \phi_1 S_{t-1} - \phi_2 S_{t-2}) \\ (S_t - \phi_1 S_{t-1} - \phi_2 S_{t-2})^2 - \sigma_\varepsilon^2 \end{bmatrix} = 0 \tag{9}$$

where $A_0 = 0$, $A_1 = 1$, and $A_{i+1} = \phi_1 A_i + \phi_2 A_{i-1}$ for $i = 1, \ldots, 6$. In particular, the first line in (9) describes the moments obtained from the SIPC model, and the second and third lines describe the moments obtained from the $AR(2)$ model.

[25] The $AR(2)$ model is extensively employed in the literature as a univariate benchmark model to predict inflation, e.g., in Khan and Zhu (2006) or Batchelor (1982), because it is the simplest univariate model with dynamics rich enough to represent actual inflation series.

As we can see from Table 9, which reports the results of the estimation of (9), the overall results are similar to those found in Section 3.3. Hansen's test of over-identified restrictions is rejected in all specifications but one at the 10% level. In general, the point estimates of λ_T vary significantly according to the specification chosen, suggesting that the estimation algorithm is not robust to small sample biases. As in Section 3.3, I re-estimate the o.c. (9) twice, first discarding the o.c. related to the conditional variance and estimating (9) for $i = 1, \ldots, 6$, and then discarding the o.c. related to the lagged covariances and estimating (9) for $i = 0$. Results are reported in Table 10. Now, all of the specifications are accepted by the data − the J-statistic ranges in [0.13, 0.94] − and the pattern of the point estimates λ_T resembles that of λ_T^{2s}, with higher estimates of information duration, i.e., around 6 months, when using the SIPC model to match the lagged covariance and significantly lower estimates otherwise, i.e., at around 4 months. MR's calibration of $\lambda = 0.25$ is rejected in all specifications. Also, λ_T appears slightly more efficient than λ_T^{2s}, with standard errors around $15 - 20\%$ lower.

4.3 Theoretical robustness: a Hybrid SIPC model

The SIPC model has been criticized in the literature because it predicts that each firm changes its price in every period, even though robust evidence has been provided − across countries and sample periods − that many firms leave prices unchanged for long periods.[26]

This criticism is particularly interesting in view of our results because it could explain why the SIPC model predicts inflation persistence that is lower than the actual value when used to match the conditional variance of inflation. Intuitively, in a model with both adaptive − that is, firms that do not change the price − and inattentive producers, the effect of a shock on inflation persists over time for two reasons. First, because of the behavior of inattentive agents behavior who use past information to set future prices; second, because of adaptive producers who use lagged prices to set their current prices. Hence, the covariance between current inflation and lagged shocks in such a model would depend both on the frequency of information updating and on the size of the fraction of adaptive producers. In particular, persistence may be high even with a low degree of information stickiness insofar as there is a large fraction of adaptive producers. In this case, if we estimated the misspecified SIPC model to match inflation persistence, it is reasonable to expect that we would find downward-biased estimates of λ_T^{2s}.[27]

Our previous intuition is supported by the results from a simple simulation exercise. I compute the variance and the first order autocorrelation function of fitted inflation using three different models as data generating processes: (i) the original SIPC model; (ii) a hybrid version of the SIPC model, which will be described hereafter, where 40% of producers in the economy are adaptive firms; (iii) the same hybrid SIPC model with 80% of adaptive firms. Figure 2 reports the variance and acf(1) plotted for all values of $\lambda \in (0, 1]$ used to generate fitted inflation from the three models. The figure indicates that the introduction of adaptive firms does not significantly affect the variance of fitted inflation for any value of

[26]See Angeloni-Aucremanne-Ehrmann-Gali-Levin-Smets (2006).

[27]This follows immediately from (3). If inflation persistence increases because of adaptive agents but structural λ remains constant, then in the misspecified o.c. (3) the RHS increases, and estimated λ_T^{2s} must decrease in order for the equality to hold.

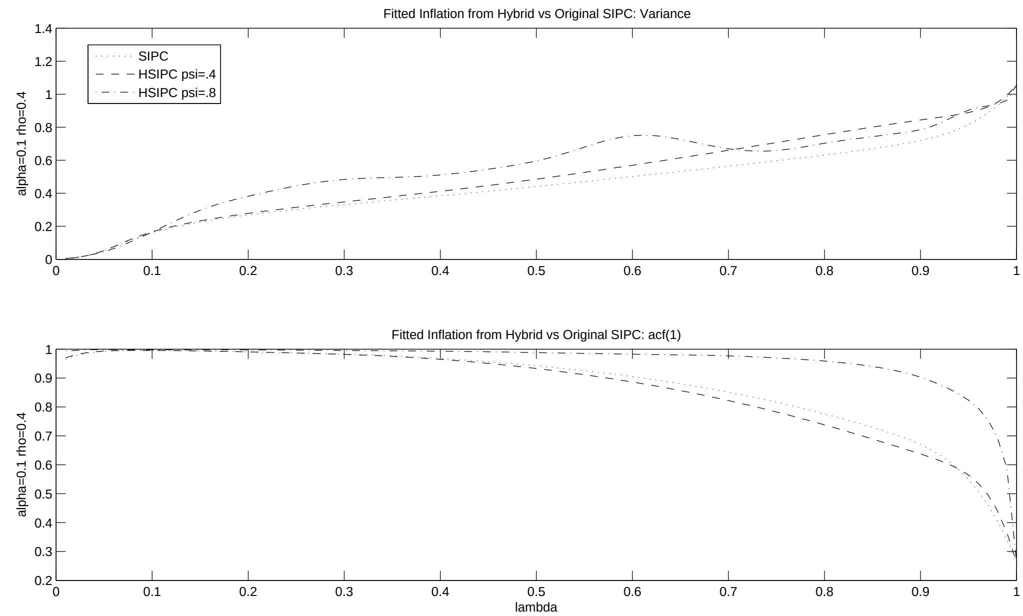

Figure 2: **Hybrid vs Original SIPC model: simulated moments of fitted inflation.** Fitted Inflation is generated either from the hybrid (HSIPC) or the original SIPC model, varying the sticky information parameter $\lambda \in (0, 1]$. Plot 1 refers to the variance of inflation, and plot 2 to the first-order autocorrelation function.

λ, but it does significantly affect persistence. The strongest effect is observed in the model with the largest fraction of adaptive firms and low values of information stickiness ($\varphi = 0.8$ and $\lambda > 0.5$). Now, if this last model were the data generating process but we estimated the original and now mispecified SIPC model, it is likely that we would find a downward bias in λ_T^{2s} when using the SIPC model to match inflation persistence because we would force the model to address all the observed persistence generated by various sources to sticky information.

In order to test the prediction of previous simulation, I derive and estimate a hybrid model with heterogeneous − inattentive and adaptive − producers.[28] Specifically, the economy is composed of two types of producers. The first type consists of a fraction φ of inattentive firms that receive information sporadically as in the SIPC model. The other $(1 - \varphi)$ fraction consists of purely adaptive firms. Following what is usually done in the literature for similar models, I use two alternative assumptions to model adaptive agents. Either they set their current price equal to last period price, as in Gali and Gertler (1999);

[28]Dupor-Kitamura-Tsuruga (2006) developed a similar model. They proposed a model of "dual stickiness" where producers change prices sporadically and absorb the relevant information for price setting in random periods, as in the SIPC. Basically, Dupor-Kitamura-Tsuruga nested together Calvo's sticky price framework with the sticky information of MR. As a result, inflation in period t is function of all past periods' expectations of future variables indexed from $t + 1$ onwards. The econometric strategy presented in Section 3.1 is computationally burdensome when applied to such a model and, therefore, in this paper I don't estimate Dupor-Kitamura-Tsuruga' model but I rather derive and estimate an alternative model with sticky information and adaptive agents.

or they index their current price to last period's price adding the latest observed inflation, as in Christiano-Eichenbaum-Evans (2005). As a result, the aggregate price index is given by

$$p_t = \varphi p_t^{SI} + (1-\varphi)\, p_t^b$$

where

$$\begin{aligned} p_t^{SI} &= \lambda \sum_{j=0}^{\infty} (1-\lambda)^j E_{t-j}(p_t + \alpha y_t) \\ p_t^b &= \begin{cases} p_{t-1} & (b_1) \\ p_{t-1} + \pi_{t-1} & (b_2) \end{cases} \end{aligned}$$

Similar to Mankiw and Reis (2002), it can be shown that inflation dynamics in this hybrid model evolves according to

$$\pi_t = \alpha\phi y_t + (1-\lambda)\,\phi \sum_{j=0}^{\infty} (1-\lambda)^j E_{t-j-1}(\pi_t + \alpha\Delta y_t) + (1-\lambda)\,\kappa\pi_{t-1} \tag{10}$$

$$\pi_t = \alpha\phi y_t + (1-\lambda)\,\phi \sum_{j=0}^{\infty} (1-\lambda)^j E_{t-j-1}(\pi_t + \alpha\Delta y_t) + \kappa\pi_{t-1} + (1-\lambda)\,\kappa\Delta\pi_{t-1} \tag{11}$$

where $\phi = \dfrac{\lambda\varphi}{1-\lambda\varphi}$ and $\kappa = \frac{1-\varphi}{1-\lambda\varphi}$. Equations (10) and (11) refer respectively to models (b_1) and (b_2).

Following the same econometric procedure presented in Section 3.1, parameters from the two models (10) and (11) can be estimated with the GMM using the following sets of orthogonality conditions:

$$\begin{aligned} E\left[\left(\frac{\alpha\lambda}{1-\lambda} y_t + \alpha\Delta y_t - \frac{1-\varphi}{\varphi(1-\lambda)}\pi_t + \frac{1-\varphi}{\varphi}\pi_{t-1}\right)(\delta\varepsilon_{t-i})'\right] &= \qquad (12) \\ &= (1-\lambda)^i \delta A_i \Sigma \delta' \\ E\left[\left(\frac{\alpha\lambda}{1-\lambda} y_t + \alpha\Delta y_t - \frac{1-\varphi}{\varphi(1-\lambda)}\Delta\pi_{t-1} + \frac{1-\varphi}{\varphi}\Delta\pi_{t-2}\right)(\delta\varepsilon_{t-i})'\right] &= \qquad (13) \\ &= (1-\lambda)^i \delta A_i \Sigma \delta' \\ for \quad i &= 0, ..., L \end{aligned}$$

where (12) and (13) refer respectively to models (b_1) and (b_2). Analogously to previous estimations, I substitute the regressors $\{\varepsilon_t, A_i, \Sigma\}$ with $\left\{\widehat{\varepsilon}_t, \widehat{A}_i, \Sigma_T\right\}$ obtained from the estimates of the VAR(p) model in Section 3.2.[29]

The results for the estimation of the o.c., (12) and (13), are reported in Table 8. The most relevant difference with the estimation of the o.c. (3) is the acceptance of the null hypothesis of over-identifying restrictions in Hansen's test. However, several point estimates

[29]The reduced form VAR(p) model used in section 3.2 encompasses both the SIPC model and the hybrid versions derived in this section.

of both λ_T^{2s} and φ_T^{2s} lie outside of the theoretical upper bounds of these parameters, raising concern about the reliability of these results. In conclusion, the validation test of the sticky information hypothesis as an explanation of inflation dynamics appears once again to be negative, and the hypothesis of adaptive producers seems not to be the key for improving the empirical performance of the theory.

5 An application: the Great Moderation

During the past 20 years, the U.S. economy has experienced a process of disinflation accompanied by a fall in inflation volatility. Some authors pointed out that during this period, known in the macroeconomic literature as the "Great Moderation", there has been not only a fall of inflation variance but also a reduction of inflation persistence, as recently confirmed by Cogley-Primicieri-Sargent (2010). Bayoumi and Sgherri (2004) pointed out that the New Keynesian models with nominal rigidities, which have been used to rationalize and explain the causes of the Great Moderation, are usually not consistent with a contemporaneous reduction in the level, variance, *and* persistence of inflation. Therefore, they concluded that the reduction of inflation persistence during the Great Moderation was due to reasons other than nominal rigidities.

In this section, I test whether there was a reduction of information stickiness in the last 20 years of the postwar sample. If we believe that firms' average information duration decreased in the postwar sample, given the development of new technologies for the transmission of information in the past 50 years, then we should expect that the contribution of sticky information to actual inflation persistence diminished during our sample period, explaining fully or partially the reduction in inflation persistence observed during the Great Moderation.

Table 4: **Test of structural breaks**

Andrews's test		sup LM	Asymptotic critical values		
H_0 : No structural breaks		statistics	10%	5%	1%
O.C. (3)	γ_0= .2	1.22	6.80	8.45	11.69
$i = 1, ..., 6$	γ_0= .1	1.21	7.63	9.31	12.69
O.C. (7)	γ_0= .2	2.08	6.80	8.45	11.69
$i = 0$	γ_0= .1	1.84	7.63	9.31	12.69

γ_0 indicates the percent of each tail of the data sample cut. SupLM statistics has non-standard distribution. The asymptotic critical values are given in Andrews (1993).

In the SIPC model, an increase in the frequency of information updating coincides with an increase in λ during the sample. I check for the presence of possible structural breaks in parameter λ using Andrews' (1993) test, which tests for structural breaks with unknown timing during the sample.[30] Andrews' test is applied to the GMM estimations presented

[30] Andrews' test cuts the tails of the data sample and then computes recursively for each observation in the middle subsample the most likely point in time where a break might have occurred.

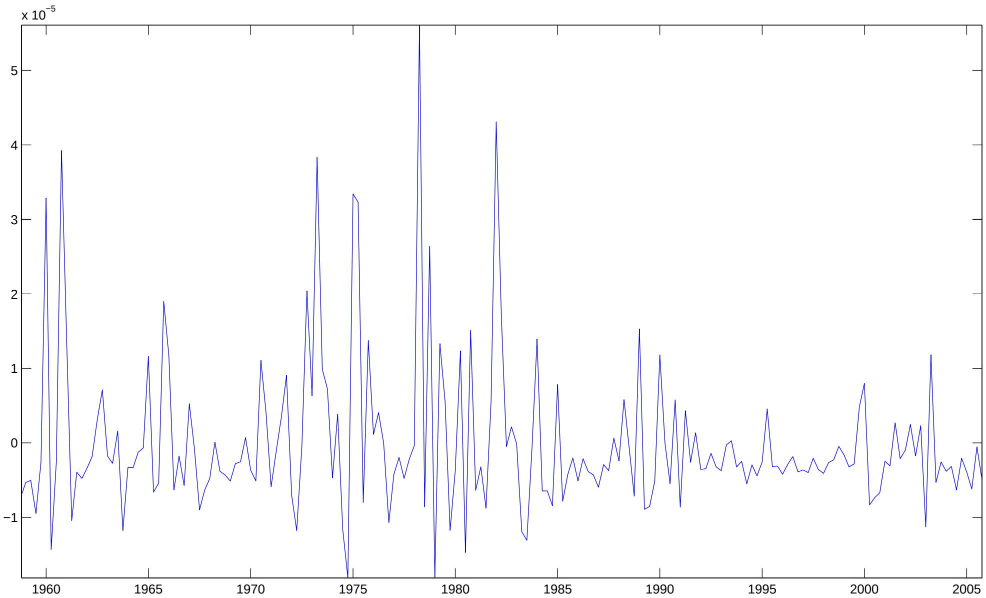

Figure 3: **Estimated GMM residuals over the sample**.

in Section 3.3, and results are reported in Table 4. According to this test, *no* structural break in λ occurred during the U.S. postwar data sample or, equivalently, the frequency of information updating by firms did not increase in the past 50 years. In our analysis, this evidence implies that the contribution of sticky information to inflation persistence remained constant during the sampled period, thus discarding the hypothesis that sticky information was a key determinant of the reduction of inflation persistence during the Great Moderation.

Although unambiguous, the result of the test leaves some reason of concern about a possible false-positive result (Type II error) due to the application of Andrews test to the particular historical period considered. During the late 1970s, inflation volatility sharply increased because of the oil shock. Although this was undoubtedly an exogenous event with respect to the objective of our analysis, it surely affected the second order moments of the inflation series, which is the information from the data exploited in our estimations of λ. In particular, because (i) Andrews test detects as structural break any significant difference in supLM statistics calculated in the first and last parts of the sample and (ii) the oil shock occurred in the middle of our sample and lasted long enough to be always included in the subsample supLM statistics, it is possible that the contribution of the oil shock to the magnitude of the second-order moments of the inflation process was too large for Andrews test to detect the effects of changes in the degree of information stickiness.

A sense of the impact of the oil shock in our estimations is given in the next Figure (3), which plots the residuals from the estimation of the o.c. (7) over the sample.[31] The large effect of the oil shock on the conditional variance is apparent in the middle of the sample,

[31] The Figure for the estimation of o.c. (3) is very similar and it is available upon request.

suggesting that it is important to control for the effect of the oil shock before testing for structural breaks in the sticky information parameter λ. To this end, I assume that the same SIPC model holds throughout the sample, and then I test whether λ is equal in two subsamples that do not include the oil shock period − that goes from 1958q4 to 1972q3, and one from 1988q3 to 2005q4 −. Subsample estimates can then be used to compare information stickiness in the 1960s versus the 1990s. Table 5 reports the subsample estimates and the results of the statistics − Wald and LM − used to test the null hypothesis that $\lambda_{60}^{2s} = \lambda_{90}^{2s}$. The table also reports the estimated information stickiness in the mid-subsample, i.e., the portion that includes the oil shock period. This is done to check whether λ_T^{2s} spuriously adapt in response to the high inflation experienced by the economy at that time.

Table 5: **Test of structural breaks: equal λ_T^{2s} in different subsamples**

Restricted $\alpha = .2$	λ_{60}^{2s}	λ_{70}^{2s}	λ_{90}^{2s}	$Wald$	LM
$H_0 : \lambda_{60}^{2s} = \lambda_{90}^{2s}$	(s.e.)	(s.e.)	(s.e.)	(p-val)	(p-val)
O.C. (3)	0.40	0.32	0.51	2.16	26.51
$i = 1, ..., 6$	(0.068)	(0.143)	(0.08)	(0.14)	(0.00)
O.C. (7)	0.48	0.40	0.92	79.94	1971.1
$i = 0$	(0.054)	(0.055)	(0.035)	(0.00)	(0.00)

Estimates of λ_{60}^{2s}, λ_{70}^{2s}, and λ_{90}^{2s} refer to data samples, respectively, from 1958q4 to 1972q3, from 1971q2 to 1982q2, and from 1988q3 to 2005q4. The Null Hypothesis for Wald and LM tests is $\lambda_{60}^{2s} = \lambda_{90}^{2s}$. Both tests have standard $\chi^2(1)$ distribution.

In general, the point estimates of λ_T^{2s} differ significantly among subsamples. First and most importantly, when using either the o.c. (3) or the o.c. (7), λ_{60}^{2s} appears lower than λ_{90}^{2s}. Also, estimated λ_{70}^{2s} is always lower than λ_{60}^{2s} and λ_{90}^{2s}, suggesting that λ_{70}^{2s} is in fact downward biased due to the oil shock. Finally, it worth noting that the standard errors of λ_{90}^{2s} are lower than those of λ_{60}^{2s} when using the contemporaneous covariance as o.c., reasonably because inflation was sensibly less volatile during the Great Moderation of late eighties and nineties than in previous periods.

Regarding the test of equal values of λ in different subsamples, both the Wald and the LM tests reject the null hypothesis for all of the specifications but one. According to this result, it is likely that the frequency of information updating actually changed between the 1960s and the 1990s. As mentioned before, this result may exist for several reasons: e.g., more media to channel macroeconomic news, more accurate forecasts about market conditions, more experienced authorities that release the relevant information, etc. It is not surprising, then, that firms acquired information more often in the 1990s than in the 1960s, thus taking less time to react to new events, which in turn made inflation less persistent.

6 Conclusions

The estimates of firms' average information duration provided in this paper show that the SIPC model can explain two crucial moments of actual inflation process − persistence and variance − only assuming two different extents of information stickiness in the economy. In detail, when the SIPC model is estimated by matching the covariances between current inflation and lagged shocks, then $\lambda_T^{2s} \in [0.35, 0.57]$. This value implies an average information duration of 6 to 9 months, in line with previous estimates that used single equation estimators to estimate the SIPC. In contrast, when the SIPC model is estimated by matching the covariances between current inflation and contemporaneous shocks, then $\lambda_T^{2s} \in [0.71, 0.86]$. This value implies an average firm's information duration of 3.5 to 4 months, which significantly differs from previous range, and turns out to be just slightly longer than the average information duration in the neoclassical model with rational expectations and flexible prices.

As showed in Section 4, previous finding appears robust to several empirical tests and to some theoretical deviations from the original SIPC model. In particular, a simulation exercise proposed in Section 4.3 suggests that the estimates of firms' average information duration may be upward biased when in fact there are multiple source of rigidities in firms' prices settings. I investigate this issue introducing adaptive firms as second source of inflation persistence. This modification of the original SIPC model, however, does not appear to reconcile the sticky information theory with the data because the estimated model with sticky information and adaptive firms is accepted by the data only for estimated values of parameters that lie outside the bounds imposed by the theory.

Finally, by analyzing sticky information in different subsamples, I find evidence that firms' average information duration was significantly longer in the first years of the sample (1960s) than in the last ones (1990s). This finding suggests that: (i) sticky information might have been an important source of inflation persistence in past times, but this seems not to be the case in recent times; (ii) part of the reduction of inflation persistence observed during the Great Moderation may depend on the smaller contribution of sticky information as source of inflation persistence.

References

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

[2] Angeloni, Ignazio and Luc Aucremanne and Michael Ehrmann and Jordi Galí and Andrew Levin and Frank Smets (2006): "New Evidence on Inflation Persistence and Price Stickiness in the Euro Area: Implications for Macro Modeling," Journal of the European Economic Association, vol. 4 n.2-3, pp. 562-574

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

[4] Christiano, Lawrence and Martin Eichenbaum and Charles Evans (2005): "Nominal rigidities and the dynamic effects of a shock to monetary policy" Journal of Political Economy, vol. 113 n. 1, pp. 1-45

[5] Cogley, Timothy, Giorgio E. Primicieri and Thomas J. Sargent (2010): "Inflation-Gap Persistence in the US", American Economic Journal: Macroeconomics, vol. 2:1 pp. 43-69

[6] Coibion, Olivier (2010): "Testing the Sticky Information Phillips Curve", The Review of Economics and Statistics vol. 92, pp. 87-101

[7] Dupor, Bill and Tomiyuki Kitamura and Takayuki Tsuruga (2005): "Do sticky Prices Need to Be Replaced with Sticky Information?", IMES Discussion Paper Series n. 2006-E-23, Bank of Japan

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

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

[10] Khan, Hashmat and Zhenhua Zhu (2006): "Estimates of the Sticky-Information Phillips Curve for the United States", Journal of Money, Credit, and Banking vol. 38 n. 1, pp. 195-207

[11] Kiley, Michael T. (2006): "A Quantitative Comparison of Sticky-Price and Sticky-Information models of price setting", Finance and Economics Discussion series n. 2006-45, Federal Reserve Board

[12] Korenok, Oleg (2008): "Empirical comparison of sticky price and sticky information models", Journal of Macroeconomics, vol. 30(3), pp. 906-927

[13] Laforte, Jean-Philippe (2007): "Pricing Models: A Bayesian DSGE approach for the US", Journal of Money Credit and Banking vol. 39, pp. 127-154

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

[15] Mankiw, Gregory N. and Ricardo Reis (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

[16] Mankiw, Gregory N. and Ricardo Reis (2007): "Sticky Information in General Equilibrium", Journal of the European Economic Association vol 5, pp. 603-613

[17] Molinari, Benedetto (2007): "The role of Sticky Information in Inflation Dynamics: Estimates and Findings", Ente Luigi Einaudi Temi di Ricerca n. 50

[18] Reis, Ricardo (2006): "Inattentive Producers", Review of Economic Studies vol. 73, pp.793-821

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

[20] Wang, Pengfei and Yi Wen (2006): "Solving Linear Difference Systems with Lagged Expectations by a Method of Undetermined Coefficients", Federal Reserve Bank of St. Louis, Working Paper 2006-003C.

[21] Woodford, Michael (2003): "Interest and Prices", Princeton University Press

A Proof of Proposition 1

Denoting the j-periods-ahead forecast error $\varepsilon^F_{t|t-j} = Z_t - E\left[Z_t \mid \Omega_{t-j}\right]$, the Sticky Information Phillips Curve can be written as:

$$\pi_t = \frac{\alpha\lambda}{1-\lambda} y_t + \frac{\lambda}{1-(1-\lambda)} \delta Z_t - \lambda \sum_{j=0}^{\infty} (1-\lambda)^j \, \delta \varepsilon^F_{t|t-j-1} \tag{14}$$

where Z_t is any covariance stationary vector of variables that includes inflation and output gap, and δ is a $(1 \times n)$ row vector of zeros and constants that picks $(\pi_t + \alpha \Delta y_t)$ within Z_t. From equation (14) and from the definition of δZ_t it is immediate to see that:

$$\frac{\alpha\lambda}{1-\lambda} y_t + \alpha \Delta y_t = \lambda \sum_{j=0}^{\infty} (1-\lambda)^j \, \delta \varepsilon^F_{t|t-j-1} \tag{15}$$

Now, using the Wold decomposition of Z_t,

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

we have:

$$\varepsilon^F_{t|t-j-1} = \sum_{i=0}^{j} A_i \varepsilon_{t-i} \tag{16}$$

and using (16) to substitute out $\varepsilon^F_{t|t-j-1}$ in the RHS of (15) we obtain:

$$\frac{\alpha\lambda}{1-\lambda} y_t + \alpha \Delta y_t = \lambda \sum_{j=0}^{\infty} (1-\lambda)^j \, \delta \sum_{i=0}^{j} A_i \varepsilon_{t-i} \tag{17}$$

Moreover, notice that:

$$\sum_{j=0}^{\infty} (1-\lambda)^j \, \delta \sum_{i=0}^{j} A_i \varepsilon_{t-i} =$$
$$\left(\delta\varepsilon_t + (1-\lambda)\,\delta\varepsilon_t + (1-\lambda)^2\,\delta\varepsilon_t + \ldots\right) + \left((1-\lambda)\,\delta A_1 \varepsilon_{t-1} + (1-\lambda)^2\,\delta A_1 \varepsilon_{t-1} + \ldots\right) + \ldots =$$
$$\frac{1}{1-(1-\lambda)} \sum_{i=0}^{\infty} (1-\lambda)^i \, \delta A_i \varepsilon_{t-i} \tag{18}$$

Hence, plugging (18) into (17) 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 Proposition.

B Proof of Proposition 2

Let $g_{1,t}$ be the set of orthogonality conditions (3), i.e.

$$g_{1,t} \equiv \underset{m\times 1}{g_1(\lambda,\beta,Y_t)} = \left[\begin{array}{c} \left(\frac{\alpha\lambda}{1-\lambda}y_t + \alpha\Delta y_t\right)\delta\varepsilon_t - (1-\lambda)\delta A_0\Sigma\delta' \\ \cdots \\ \left(\frac{\alpha\lambda}{1-\lambda}y_t + \alpha\Delta y_t\right)\delta\varepsilon_{t-L} - (1-\lambda)^L\delta A_L\Sigma\delta' \end{array}\right]$$

where $m = L+1$ is the number of orthogonality conditions, $\beta = vec\left(\left[B_1', \ldots, B_p'\right]'\right)$, B_j are the matrices of parameters defined in (5), $vec(\cdot)$ is the column stacking operator, and $Y_t = \{y_t, y_{t-1}\}$. As showed in (3), the SIPC model implies that:

$$E\left[g_1\left(\lambda,\beta,Y_t\right)\right] = 0 \tag{19}$$

Also, let $g_{2,t}$ be the orthogonality conditions used to estimate the VAR(p) model (5), therefore holding

$$E\left[\underset{kn\times 1}{g_2\left(\beta,X_t\right)}\right] = 0 \tag{20}$$

where X_t are the n endogenous variables of the VAR(p), and $k = np+1$. [32]

First step: One-stage GMM estimator of $\{\beta,\lambda\}$

The 2-stages GMM estimator λ_T^{2s} provided in section 3.1 estimates λ using the following o.c.

$$E\left[g_1\left(\lambda,\beta_T^{VAR},Y_t\right)\right] = 0$$

where β_T^{VAR} are the estimated coefficients of the VAR(p) model (20). It can be shown that this is akin to estimate jointly $\{\lambda,\ \beta\}$ with a GMM estimator that uses the pooled vector of o.c.

$$E\left[\begin{array}{c} \underset{m\times 1}{g_{1,t}} \\ \underset{kn\times 1}{g_{2,t}} \end{array}\right] = 0 \tag{21}$$

where we stack column-wise (19) and (20), and a weighting matrix W that satisfies:

$$W = \left(\begin{array}{cc} \Sigma_{g_1}^{-1} & O \\ O' & I_{kn} \end{array}\right) \tag{22}$$

Denote the one-stage estimator of λ from the pooled model (21) as λ_T. It is immediate to show that the point estimates of the two estimators λ_T^{2s} and λ_T coincides, once noting that by construction $\dfrac{\partial E\left[g_2\left(\beta,X_t\right)\right]}{\partial\lambda} = 0$ and that W is block diagonal. In this case, the objective function of the one-stage GMM estimator of the pooled model (21) coincides with

[32] As usual in the literature on VAR, the VAR(p) model (5) is assumed to have errors $\varepsilon_t \sim i.i.d.\ N(0,\Sigma)$, and it is estimated LS equation by equation.

the objective function of the two-stage GMM estimator λ_T^{2s} when the optimal weighting matrix is used, i.e. $J(\lambda_T) = E(g_{1,t}) \cdot W_{1,1} \cdot E(g'_{1,t})$ with $W_{1,1} = \Sigma_{g_1}^{-1}$.
Notice, however, that the variance of the two estimators are different. The reason is that the not-adjusted variance of λ_T^{2s} computed by the GMM algorithm,

$$V_{na}(\lambda_T^{2s}) = \left(TE\frac{\partial g'_{1,t}}{\partial \lambda} \Sigma_{g_1}^{-1} E\frac{\partial g_{1,t}}{\partial \lambda} \right)^{-1} \tag{23}$$

treats β_T^{VAR} as a fixed variable and not as a stochastic regressor. On the contrary, in the pooled model (21) the effect of β on $g_{1,t}$ is directly accounted, and β_T is jointly estimated with λ_T using fixed data. Thus, the variance of λ_T correctly accounts for the effect of β_T, and I suggest to use this as the *adjusted variance* of λ_T^{2s}.

Second step: derive the variance of λ_T

The VCV matrix of λ_T, if W is the optimal weighting matrix for the model (21), is:

$$V\begin{pmatrix} \lambda_T \\ \beta_T \end{pmatrix} = \left[T \cdot \underbrace{E\left(\frac{\partial \begin{pmatrix} g_{1,t} \\ g_{2,t} \end{pmatrix}}{\partial \begin{pmatrix} \lambda & \beta' \end{pmatrix}} \right)'}_{\equiv G} \underbrace{\left(E\begin{pmatrix} g_{1,t} \\ g_{2,t} \end{pmatrix} \begin{pmatrix} g'_{1,t} & g'_{2,t} \end{pmatrix} \right)^{-1}}_{\equiv \Omega} E\left(\frac{\partial \begin{pmatrix} g_{1,t} \\ g_{2,t} \end{pmatrix}}{\partial \begin{pmatrix} \lambda & \beta' \end{pmatrix}} \right) \right]^{-1} \tag{24}$$

where G is a $(kn + m) \times (kn + 1)$ matrix. The variance (24) can be written as:

$$V\begin{pmatrix} \lambda_T \\ \beta_T \end{pmatrix} = \left[E\begin{pmatrix} \frac{\partial g'_1}{\partial \lambda} & O' \\ \frac{\partial g'_1}{\partial \beta} & \frac{\partial g'_2}{\partial \beta} \end{pmatrix} \begin{pmatrix} \Sigma_{g_1}^{-1} & O' \\ O & \Sigma_{g_2}^{-1} \end{pmatrix} E\begin{pmatrix} \frac{\partial g_1}{\partial \lambda} & \frac{\partial g_1}{\partial \beta'} \\ 0 & \frac{\partial g_2}{\partial \beta'} \end{pmatrix} \right]^{-1} /T$$

where $\Sigma_{g_h} = E\left(g_{h,t} \cdot g'_{h,t}\right)$ for $h = \{1, 2\}$ and O denotes a matrix of zeros of needed dimensions. After some algebra manipulation, this variance can be written as:

$$V\begin{pmatrix} \lambda_T \\ \beta_T \end{pmatrix} = \begin{pmatrix} E\frac{\partial g'_{1,t}}{\partial \lambda}\Sigma_{g_1}^{-1}E\frac{\partial g_{1,t}}{\partial \lambda} & E\frac{\partial g'_{1,t}}{\partial \lambda}\Sigma_{g_1}^{-1}E\frac{\partial g_{1,t}}{\partial \beta'} \\ E\frac{\partial g'_{1,t}}{\partial \beta}\Sigma_{g_1}^{-1}E\frac{\partial g_{1,t}}{\partial \lambda} & E\frac{\partial g'_{1,t}}{\partial \beta}\Sigma_{g_1}^{-1}E\frac{\partial g_{1,t}}{\partial \beta'} + E\frac{\partial g'_{2,t}}{\partial \beta}\Sigma_{g_2}^{-1}E\frac{\partial g_{2,t}}{\partial \beta'} \end{pmatrix}^{-1} /T \tag{25}$$

In particular, using the definition (23), and the simplifying formula for the inverse of partitioned matrices, the variance of λ_T in from equation (25) can be written as:

$$V(\lambda_T) = \left(\left(TV_{na}(\lambda_T^{2s})\right)^{-1} - E\frac{\partial g'_{1,t}}{\partial \lambda}\Sigma_{g_1}^{-1}E\frac{\partial g_{1,t}}{\partial \beta'} \left(E\frac{\partial g'_{1,t}}{\partial \beta}\Sigma_{g_1}^{-1}E\frac{\partial g_{1,t}}{\partial \beta'} + E\frac{\partial g'_{2,t}}{\partial \beta}\Sigma_{g_2}^{-1}E\frac{\partial g_{2,t}}{\partial \beta'} \right)^{-1} E\frac{\partial g'_{1,t}}{\partial \beta}\Sigma_{g_1}^{-1}E\frac{\partial g_{1,t}}{\partial \lambda} \right)^{-1} /T \tag{26}$$

which proves the proposition, and accordingly, it is used in the rest of the paper as the variance of λ_T^{2s} adjusted for the stochastic regressors.

Notice that previous derivation uses the assumption that the covariance between $g_{1,t}$ and $g_{2,t}$ is zero. If we relax this assumption, i.e. if $E\left(g_{1,t} \cdot g_{2,t}'\right) \neq 0$ then the variance of λ_T will be the upper left cell of:

$$V \begin{pmatrix} \lambda_T \\ \beta_T \end{pmatrix} = \left(G'WG\right)^{-1} G'W\Omega WG \left(G'WG\right)^{-1} /T \tag{27}$$

where G is defined as in equation (24), and this will be the adjusted variance of λ_T^{2s}. Notice also that in this case the two-stages estimator λ_T^{2s} is no longer the most efficient estimator of λ among the GMM estimators of (21).

C Robustness Analysis: Tables

Table 6: **Empirical Robustness, restricted** $\alpha = 0.1$

Restricted	$\alpha = .1$	λ_T^{2s} (adj. s.e.)	MR calib. $H_0 : \lambda$	t-stat (p-val)	RE calib. $H_0 : \lambda$	t-stat (p-val)	J-stat (p-val)
O.C. (3)	$i = 0, ..., 6$						
	(a)	0.93	0.25	14.03	1	-1.32	14.62
GDP deflator		(0.049)		(0.00)		(0.18)	(0.02)
baseline	(b)	0.95	0.25	15.08	1	-0.96	10.09
		(0.046)		(0.00)		(0.33)	(0.12)
O.C. (3)	$i = 1, ..., 6$						
	(a)	0.51	0.25	3.74	1	-6.77	1.91
GDP deflator		(0.071)		(0.00)		(0.00)	(0.86)
baseline	(b)	0.53	0.25	3.82	1	-6.36	1.62
		(0.074)		(0.00)		(0.00)	(0.89)
O.C. (7)	$i = 0$						
	(a)	0.94	0.25	15.98	1	-1.26	16.96
GDP deflator		(0.043)		(0.00)		(0.20)	(0.52)
baseline	(b)	0.95	0.25	16.09	1	-1.03	13.37
		(0.044)		(0.00)		(0.30)	(0.76)

Estimates of λ_T^{2s} are obtained from 2-stages GMM with optimal weighting matrix and o.c. as indicated in column 1. Data sample 1958q4 – 2005q4. Output gap is GDP filtered with HP filter. *adj. s.e.* are Newey-West HAC standard errors adjusted for stochastic regressors. p-values in parenthesis. *J-stat* is Hansen's test of overidentifying restrictions (with 5 d.o.f. for o.c. (3), and 18 d.o.f. for o.c. (7)).

Table 7: **Empirical Robustness, QD filter**

Restricted	$\alpha = .2$	λ_T^{2s}	MR calib.	t-stat	RE calib.	t-stat	J-stat
		(adj. s.e.)	$H_0 : \lambda$	(p-val)	$H_0 : \lambda$	(p-val)	(p-val)
O.C. (3)	$i = 0, ..., 6$						
	(a)	0.33	0.25	1.30	1	-10.19	23.49
GDP deflator		(0.065)		(0.19)		(0.00)	(0.00)
baseline	(b)	0.86	0.25	5.56	1	-1.21	13.26
		(0.110)		(0.00)		(0.22)	(0.03)
O.C. (3)	$i = 1, ..., 6$						
	(a)	0.29	0.25	0.83	1	-6.77	1.35
GDP deflator		(0.056)		(0.40)		(0.00)	(0.92)
baseline	(b)	0.30	0.25	0.91	1	-12.09	1.17
		(0.058)		(0.36)		(0.00)	(0.94)
O.C. (7)	$i = 0$						
	(a)	0.61	0.25	4.30	1	-4.55	33.07
GDP deflator		(0.085)		(0.00)		(0.00)	(0.02)
baseline	(b)	0.79	0.25	5.67	1	-2.14	11.15
		(0.096)		(0.00)		(0.03)	(0.88)

Estimates of λ_T^{2s} are obtained from 2-stages GMM with optimal weighting matrix and o.c. as indicated in column 1. Data sample 1958q4 – 2005q4. Output gap is GDP filtered with Quadratic Detrend filter. *adj. s.e.* are Newey-West HAC standard errors adjusted for stochastic regressors. p-values in parenthesis. *J-stat* is Hansen's test of overidentifying restrictions (with 5 d.o.f. for o.c. (3), and 18 d.o.f. for o.c. (7)).

Table 8: **Theoretical Robustness: the Hybrid SIPC model**

Restricted	$\alpha = .2$	O.C. (12) with $i = 0, ..., 6$			O.C. (13) with $i = 0, ..., 6$		
		λ_T^{2s} (s.e.)	φ_T^{2s} (s.e.)	J-stat (p-val)	λ_T^{2s} (s.e.)	φ_T^{2s} (s.e.)	J-stat (p-val)
	(a)	1.14	0.64	2.14	0.48	1.21	8.63
GDP deflator		(0.120)	(0.162)	(0.82)	(0.074)	(0.094)	(0.12)
baseline	(b)	0.30	1.77	3.09	0.48	1.27	6.90
		(0.059)	(0.400)	(0.68)	(0.073)	(0.106)	(0.22)
	(a)	1.16	0.64	1.72	0.49	1.20	3.93
GDP deflator		(0.160)	(0.223)	(0.88)	(0.097)	(0.101)	(0.55)
minRMSE	(b)	0.34	1.78	1.42	0.49	1.24	3.50
		(0.090)	(0.447)	(0.92)	(0.097)	(0.111)	(0.62)
	(a)	1.08	0.75	2.92	0.63	1.17	12.82
CPI		(0.077)	(0.098)	(0.71)	(0.078)	(0.061)	(0.02)
baseline	(b)	0.36	1.57	9.11	0.63	1.21	11.14
		(0.065)	(0.27)	(0.10)	(0.077)	(0.072)	(0.04)
	(a)	1.04	0.78	1.66	0.75	1.08	6.98
CPI		(0.093)	(0.120)	(0.89)	(0.101)	(0.063)	(0.22)
minRMSE	(b)	0.50	1.40	7.73	0.74	1.10	6.98
		(0.116)	(0.24)	(0.17)	(0.100)	(0.068)	(0.22)

Estimates in the table are obtained from 2-stages GMM with optimal weighting matrix and o.c. as indicated in row 1 column 2 and 3. Data sample 1958q4 – 2005q4. HP filter for output gap. Newey-West HAC standard errors adjusted for stochastic regressors. p-values in parenthesis. *J-stat* is Hansen's test of overidentifying restrictions (with 4 d.o.f.).

Table 9: **Methodological Robustness: the pooled model estimator**

O.C. (9) Restricted	$i = 0, ..., 6$ $\alpha = .2$	λ_T (s.e.)	ϕ_1 (s.e.)	ϕ_2 (s.e.)	σ_ε^2 (s.e.)	$H_0 : \lambda = .25$ (p-val)	J-stat (p-val)
	(a)	0.90	0.67	0.28	1.2e-5	7.00	11.24
GDP deflator		(.093)	(.062)	(.065)	(0.2e-5)	(0.00)	(0.08)*
$\pi_t + \alpha\Delta y_t \sim$AR(2)	(b)	0.24	0.69	0.27	0.2e-5	-0.07	23.88
		(.030)	(.063)	(.067)	(0.1e-5)	(0.94)*	(0.00)
	(a)	0.70	0.62	0.32	1.3e-5	5.84	17.72
CPI		(.077)	(.047)	(.044)	(0.3e-5)	(0.00)	(0.00)
$\pi_t + \alpha\Delta y_t \sim$AR(2)	(b)	0.83	0.60	0.32	2.0e-5	9.97	9.04
		(.058)	(.046)	(.046)	(0.3e-5)	(0.00)	(0.17)*

Estimates in the table are obtained from 2-stages GMM with optimal weighting matrix and o.c. as indicated in row 1. The inattentive firms' forecasting technology is an $AR(2)$ model. Data sample 1958q4 – 2005q4. Output gap is GDP filtered with HP filter. Newey-West HAC standard errors (no stochastic regressors). p-values in parenthesis. *J-stat* is Hansen's test of overidentifying restrictions (with 6 d.o.f.).

Table 10: **Methodological Robustness: the pooled model estimator**

O.C. (9)	$i = 1, ..., 6$	λ_T	ϕ_1	ϕ_2	σ^2_ε	H_0: $\lambda = .25$	J-stat
Restricted	$\alpha = .2$	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(p-val)	(p-val)
	(a)	0.51	0.61	0.33	1.4e-5	3.92	0.97
GDP deflator		(.067)	(.061)	(.064)	(0.2e-5)	(0.00)	(0.96)*
$\pi_t + \alpha\Delta y_t$ ~AR(2)	(b)	0.52	0.62	0.32	1.4e-5	4.02	0.83
		(.068)	(.060)	(.063)	(0.2e-5)	(0.00)	(0.97)*
	(a)	0.54	0.66	0.28	1.9e-5	6.74	2.27
CPI		(.043)	(.051)	(.048)	(0.3e-5)	(0.00)	(0.80)*
$\pi_t + \alpha\Delta y_t$ ~AR(2)	(b)	0.55	0.64	0.29	1.9e-5	6.35	1.86
		(.058)	(.058)	(.048)	(0.3e-5)	(0.00)	(0.86)*
O.C. (9)	$i = 0$						
Restricted	$\alpha = .2$						
	(a)	0.91	0.61	0.34	1.4e-5	7.66	exactly
GDP deflator		(.086)	(.066)	(.074)	(0.2e-5)	(0.00)	identif.
$\pi_t + \alpha\Delta y_t$ ~AR(2)	(a) with	0.96	0.58	0.38	1.0e-5	9.75	24.72
	instr.	(.072)	(.044)	(.048)	(0.2e-5)	(0.00)	(0.13)*
	(a)	0.81	0.66	0.28	2.1e-5	9.59	exactly
CPI		(.058)	(.060)	(.056)	(0.4e-5)	(0.00)	identif.
$\pi_t + \alpha\Delta y_t$ ~AR(2)	(a) with	0.98	0.64	0.31	0.9e-5	10.08	24.69
	instr.	(.072)	(.038)	(.037)	(0.2e-5)	(0.00)	(0.13)*

Estimates in the table are obtained from 2-stages GMM with optimal weighting matrix and o.c. as indicated in row 1. The inattentive firms' forecasting technology is an $AR(2)$ model. Data sample 1958q4 – 2005q4. Output gap is GDP filtered with HP filter. Newey-West HAC standard errors (no stochastic regressors). p-values in parenthesis. *J-stat* is Hansen's test of overidentifying restrictions (with 5 d.o.f.).