Sticky Information and Inflation Persistence: Evidences from U.S. data

Benedetto Molinari*
Universitat Pompeu Fabra†

September 2007

Abstract

This paper provides a simulated moments estimator of the Sticky Information Phillips Curve. A crucial feature of this model is that inflation is a persistent process, i.e. the effect of a shock on inflation lasts in time because firms acquire information sporadically. Therefore, I propose to estimate the degree of information stickiness in the economy using the model to match the covariances between actual inflation and exogenous shocks.

The paper provides estimates for U.S. postwar economy. The main result is that the SIPC model does not fit the data. This finding is very different to what has been found in the literature until now. I show that this difference depends on how much information about the inflation process we use. In particular, when the model is estimated matching only the covariances between inflation and lagged shocks, then the estimates of firms' frequency of information updating λ are in accordance with that of the other papers that estimated the SIPC, i.e. $\lambda_T \in [.35, .57]$.

Differently, when the model is estimated matching the conditional variance of inflation alone, the estimates of λ are significantly higher than before, i.e. $\lambda_T \in [.71, .86]$.

The distance between these two ranges is crucial for the implications of the model. Indeed, for λ close to 1 the SIPC predicts an inflation dynamics with little persistence, which is at odd with the data.

JEL Classification: E31, C51

Keywords: Sticky Information, Inflation Persistence, GMM estimator.

---

*I'm grateful to my advisor, Albert Marcet, for his guidance. I wish also to thank Marcello De Cecco for his advises, and Francesco Turino, Claudio Bianchi, Ricardo Nuñes, and the participants to the macroeconomic seminar at UPF for useful comments. Financial support from Ente Luigi Einaudi is gratefully acknowledged. Contact email: benedetto.molinari@upf.edu

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

1 Introduction

This paper estimates the degree of information stickiness in U.S. post-war economy. The model I refer to is the Sticky Information Phillips Curve (henceforth SIPC), proposed by Mankiw and Reis (2002, henceforth MR) as structural theory of inflation.

MR goal was to understand why actual inflation responds gradually to a large number of different shocks, as observed in post-war U.S. data. As a matter of fact, several theories of prices, such as the neoclassical model with no nominal rigidities and the New Keynesian Phillips Curve, predict an inflation dynamics far less persistent than what we observe in actual data.[1] In general, if firms maximize profits and have rational expectations, they will react to any exogenous shock adjusting their prices to the new target level as soon as they acquire information about the shock. Hence, the effect of a shock on price changes disappears rapidly, and the only source of persistence in inflation dynamics can be the one of the exogenous shocks (e.g. cost push shocks, monetary policy shocks, demand shocks).

To overcome the lack of intrinsic persistence in inflation dynamics, MR model is based on the idea that firms absorb only sporadically the information they need to choose their price plans. In those periods when information inflows are limited or absent, firms set prices based on outdated information. When a shock occurs only a fraction of firms adjust the price contemporaneously, while the others delay some periods to get aware of the shock, meanwhile relying on outdated price plans. Thus, the overall effect of a shock on changes of prices lasts in time, and inflation turns out to be a persistent process as real data suggest.

Regarding the issue of inflation persistence, the key parameter in the model is the frequency of firms' information updating λ. In fact, for high (low) values of λ the SIPC model predicts low (high) persistence of inflation. An example is useful to see this point. Following Reis (2004),[2] I simulate the SIPC assuming a simple univariate $AR(1)$ model for the exogenous shocks. In figure (1) I plot for all the values of $\lambda \in (0, 1]$ the ratio between the first autocovariance function $acf(1)$ of fitted inflation from the SIPC model over the $acf(1)$ of actual U.S. postwar inflation. As we can see from figure (1) the higher is λ, the smaller is the persistence explained by the model.

In the literature there isn't consensus on the estimates of λ. Reis (2004) suggested that $\lambda = 0.25$[3] is the best parameter for the SIPC to match the per-

[1] Actually, MR presented the SIPC as an alternative theory to the New Keynesian Phillips Curve, which was criticized because it lacks of persistence. The criticisms pointed out that: (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).

[2] I thank Ricardo Reis for making the rutine of his papers freely available on his webpage.

[3] This is the calibration originally used in MR (2002) paper. In a model where a period is equal to one quarter, $\lambda = 0.25$ implies that firms update information on average once a year.

sistence of U.S. postwar inflation. After, Khan and Zhu (2002, 2006), Kiley (2006) and Korenok (2005) estimated the model using limited information estimators, and they found $\lambda_T \in [0.15, 0.4]$. While Laforte (2006) and Mankiw and Reis (2006) estimated $\lambda_T \in [0.7, 0.85]$ using full information estimators.[4] As we can see from figure (1), between the first and the second range there is a relevant difference in the degree of actual inflation persistence the model explains.[5]

In this paper I try to reconcile these different estimates proposing a simulated moments GMM estimator that exploits the information contained in the covariance functions between current inflation and current and lagged exogenous shocks that are relevant in firms' pricing decisions. The intuition is that, if the model is true, the more firms' are inattentive to new shocks (low λ), the longer time a shock today will affect future price plans, and so the longer it will be correlated with inflation.

To implement this estimator it is worth writing the expectations terms that appear in the SIPC as functions of forecast errors, and then the forecast errors in terms of exogenous shocks.[6] Once the model is transformed in this way, it is easy to derive a set of orthogonality conditions that are based on the covariances between inflation and exogenous shocks. I show them in section 3.1. Then, these orthogonality conditions are estimated pursuing a two steps approach: first, I fit a vector autoregression (VAR) model for the exogenous shocks, which is used to calculate the covariances we need. Second, I use the simulated moments to estimate λ with the GMM.

With respect to the other papers that estimated the SIPC using limited information estimators, e.g. Khan and Zhu (2002, 2006), Kiley (2006) and Korenok (2005), this econometric strategy has two advantages: (i) since my orthogonality conditions have a finite number of terms, I avoid the infinite dimensions problem usually associated with the SIPC without using any truncation or approximation of the model. (ii) I use more information about the inflation process to pin down λ_T. This last point is not straightforward to see because the other papers use different methodologies from mine. I try to make the comparison in section 2.2.

In summary, the objective of this paper is twofold: (i) to estimate the SIPC model to see whether it is a valid explanation of U.S. inflation; (ii) to show the implications of the resulting estimates of λ on inflation persistence.

The paper is organized as follows: in section 2 I review the SIPC model and the literature about sticky information and inflation persistence. In section 3 I present econometric strategy and results. Section 4 analyzes whether the degree of information stickiness changed during the sample. Some conclusions are given in section 5.

---

[4] These are "full information" estimators in the sense that the SIPC is estimated – as aggregate supply equation – joint with an equation for aggregate demand, and an equation for nominal interest rate adjustment (e.g. the Taylor rule) in a fully fledge model of the aggregate economy.

[5] Note that the SIPC with $\lambda = 1$ encompasses the RE model with monopolistic competitive firms and flexible prices.

[6] This way of writing the model is similar to that in Mankiw and Reis (2006), and ?? However, our results are contemporaneous and independent.

2 The Model

2.1 The Sticky Information Phillips Curve

The main assumption in the SIPC model is that in every period there is a fraction λ, out of a continuum $(0,1)$ of monopolistic competitive firms operating on the market, which update the information about the current economic conditions of the markets (e.g. exogenous demand shocks, changes in the nominal marginal cost, etc.). These firms charge a price p_t^* for their production in order to maximize profits conditional on the newly updated information, while the rest of firms sets their prices according to old price plans. In other words, the firms that don't updated information set a price to maximize profits conditional on outdated information. Each firm has the same probability to update information, regardless of how long has been since its last update.

Since all firms are ex-ante identical, optimal price is the same for all those firms that have information dated $t-j$, and any firm that updated its information j periods ago adjusts today price according to:

$$x_t^j = E\left[p_t^* \mid \Omega_{t-j}\right] \tag{1}$$

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

MR solved the aggregate dynamics of prices, and showed that inflation in this model evolves according to:

$$\pi_t = \frac{\alpha\lambda}{1-\lambda} y_t + \lambda \sum_{j=0}^{\infty} (1-\lambda)^j E\left[\pi_t + \alpha \Delta y_t \mid \Omega_{t-1-j}\right] \tag{2}$$

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 period t.[7]

As we can see from equation (2), in the SIPC model inflation is persistent because current inflation depends on past periods expectations about the current inflation and output growth, where past expectations are weighted with a weight that fades out at the rate $(1-\lambda)$. The mechanism is the following: suppose that in period t occurs a shock ε_t that increases the output gap. The information about this shock is included in period t information, i.e. $\varepsilon_t \subset \Omega_t$. Accordingly to (2), inflation raises contemporaneously because of the trade off term $\left(\frac{\alpha\lambda}{1-\lambda} y_t\right)$. In period $t+1$, when (2) holds for π_{t+1}, a fraction λ of agents gets aware of the shock occurred in t, so inflation raises again because $E\left[\Delta y_{t+1} \mid \Omega_{(t+1)-1-j}\right]$ is positive for $j=0$. Same happens in $t+2$, when a fraction $\lambda(1-\lambda)$ gets aware of the shock, and then in all the following periods $t+j \ for \ j>1$, when the effect of the shock on inflation fades out at rate $(1-\lambda)^j$. Hence, in this model a shock today affects future inflation level for infinite periods. This implies that the inflation process is serially correlated for many periods, as real data suggest.

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

2.2 Sticky Information and Inflation Persistence

The model of sticky information belongs to the literature about *Rational Inattention.*[8] This conjecture has been proposed as explanation of the stickiness observed in macroeconomic variables, and it is related with the early papers on limited information of Lucas (1973), Fischer (1977), Taylor (1980), Sims (1998), and Woodford (2001). MR combined some elements of Fisher's and Lucas's contributions and proposed the Sticky Information Phillips Curve to model inflation dynamics.

In their original paper, MR assumed that producers get new informations in every period with an exogenous probability λ, and they calibrated firms' average information duration of 1 year (in a model where each period is a quarter this corresponds to $\lambda = 0.25$). In that paper MR achieved their goal: fitted inflation responded gradually to several exogenous shocks like demand shock and monetary policy shock.

After the original MR work, the SIPC model has been estimated in different ways. Reis (2004) proposed a validation test based on the simulation of the model. First, he provided a rigorous microfoundation of SIPC based on cost-benefit analysis: firms gather new information only if the expected benefit of changing the price is higher than the cost of acquiring the information.[9] Then, he simulated the SIPC calibrating model parameters. Parameter α, which he found to be function of intertemporal elasticity of substitution, of Fisher elasticity of labor, and of the elasticity of demand of single-variety goods, is calibrated within the interval $\alpha \in (0.1, 0.2)$, following the RBC literature about these deep parameters. For λ Reis used the calibration $\lambda = 0.25$ originally proposed by MR. Finally, he showed that the model did a good job in matching some moments of the aggregate distribution of prices, including the first autocovariance function of inflation, which he used as measure of persistence.

Reis (2004) gave an important contribution to support the SIPC as possible explanation of inflation persistence, but there are some reasons of concern with his results. In the simulations of the SIPC he assumed an arbitrary process for exogenous shocks that is highly persistent itself. Therefore, we don't know how much persistence of fitted inflation came from the intrinsic dynamics of inflation with sticky information and how much from the exogenous shocks process.

A different approach is followed in Mankiw Reis and Wolfers (2003) that showed how the sticky information conjecture explains well the main features of expectations about inflation, as they are observed in the Michigan Consumers Survey and in the Survey of Professional Forecasters. Moreover, they showed that $\lambda = 0.25$ was the best value for the SIPC to match the moments of the distribution of inflation expectations.

Although Mankiw Reis and Wolfers surely provided an evidence to support

[8] The name *Rational Inattention* first appeared in Sims (2003).

[9] Reis (2004) is not the only paper that provide a microfoundation of inattentive agents. Branch (2004) explains individual inattentiveness as function of the increase in forecasts accuracy once new information is processed. Hence, the more new information improves the Thail index of forecasts (with respect to outdated information), the more agents are attentive.

the SIPC model to explain micro data, it is less clear whether their work can be used also to support MR calibration of λ. In fact, there is a good number of examples in the literature where macro models calibrated with micro data does not match the moments of aggregate variables distributions, which is our goal here.[10] I don't want to get into this issue here. My point is just to underline that a proper estimation of the SIPC using macro data seems of necessity to draw conclusions about the relationship between sticky information and inflation persistence.

In the last two years this has been done by several papers. Among the others, Khan and Zhu (2002, 2006), Kiley (2006) Korenok (2005), Laforte (2006), Mankiw and Reis (2006). The reason why I propose another estimation of the SIPC is mainly because the estimates of λ_T vary a lot among those papers.

Also, the estimator I pursue here improves the ones used by the other papers because it exploits more information about inflation. To show that I focus on those papers that use limited information estimators (e.g. Khan and Zhu, Coibion, Kiley) as I do here.

In general, the estimation strategies of the other papers have a common first step. They truncate the infinite sum of expectations in equation (2) at $t - j_{\max}$, and then they substitute the remaining expectations terms with the predictions of a VAR model set ad-hoc to forecast inflation and output gap. For instance, today expectations conditional on information dated $t-5$ are replaced by $proj_{t-5}(\pi_t + \alpha\Delta y_t)$.

Thus, the specification they estimate is:

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

Now, since the $proj_{t-j}(\pi_t + \alpha\Delta y_t)$ is a linear combination of lagged inflation and output gap,[11] what they do is a nonlinear regression of inflation at time t on $t - j_{\max}$ lags of inflation and output gap. In turns, this means that they exploit the covariances between inflation and lagged inflation and output gap terms.

Hence, they don't use any information regarding the covariance between current inflation and contemporaneous shocks. On the contrary, the simulated moments estimator allows us to estimate λ matching jointly the moments that measure inflation persistence, i.e. the covariances between inflation and lagged shocks, and the moments that measure the conditional variance of inflation, i.e. the covariance between current inflation and contemporaneous shocks. Clearly, if the SIPC model is the true DGP of the data we use as observables, then it

[10] One example is the Frish labor supply elasticity in the standard RBC model. The micro evidences about this elasticity point a value around 1/6, whereas to fit aggregate labor volatility the RBC model need a value of close to 1, so much higher.

[11] The sentence should read: "since the $proj_{t-j}(\pi_t + \alpha\Delta y_t)$ is a linear combination of lagged inflation and output gap, plus past values of other variables possibly included in the VAR." The specification of the VAR model differs in all the papers cited above, therefore it is not possible to make general statement about the information that come from other variables. However, the bulk of the argument remains true, since that information regards lagged variables.

should be able to match all the moments jointly, plus the estimates of λ shouldn't differ significatively matching one set of moments at the time. Unfortunately, the bulk of my results points out that this is not the case, as the reader will see in section 3.3.

3 The Estimation

3.1 Econometric strategy

I use here the standard assumption that the dynamics of inflation and output gap result from the interaction of n macroeconomic variables, which I define 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 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 (2) implies:*

$$\frac{\alpha\lambda}{1-\lambda} y_t + \alpha \Delta y_t = \sum_{i=0}^{\infty} (1-\lambda)^i \, \delta A_i \varepsilon_{t-i} \tag{3}$$

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.[12] ■

Equation (3) is useful to derive a set of orthogonality conditions. Multiplying (3) by a vector of lagged shocks ε_{t-i} for $i = 0, ..., l$ and taking the expectations, I obtain:

$$\begin{aligned} E\left[\left(\frac{\alpha\lambda}{1-\lambda} y_t + \alpha \Delta y_t\right) (\delta \varepsilon_{t-i})'\right] &= (1-\lambda)^i \, \delta A_i \Sigma \delta' \\ for \quad i &= 0, \ldots, l \end{aligned} \tag{4}$$

where $\Sigma \equiv E[\varepsilon_t \varepsilon_t']$ is the VCV matrix of the shocks.[13]

Is useful to see that each orthogonality condition in (4) matches a lag of the covariance between Z_t and a linear combination of ε_t. Also, the RHS of (4) is a linear combination of the impulse response functions (IRF) of Z_t, weighted by the frequency of firms' that don't update information. Intuitively, in the SIPC model this linear combination is function of the output gap process, which is the driving force of prices, and of λ, which measure how many firms are attentive to the shocks, i.e. how rapidly the effect of the shocks on prices fades out.

[12] This result is similar to those found by Mankiw and Reis (2006) and Wang and Wen (2006). However, I obtained it contemporaneously and independently from those papers.

[13] Equation (4) follows multiplying (3) by $(\delta \varepsilon_{t-j})'$ and taking the expectations conditional on information at time t. It uses the fact that $E[\varepsilon_t \varepsilon_{t-j}] = 0$, $j = 1, \ldots$

The moments (4) depend on the unknown regressors $\{\varepsilon_t, A_i, \Sigma\}$.[14] Hence, in order to estimate them I pursue a two steps approach. First, I estimate a vector autoregression model of Z_t to obtain consistent estimates of $\{\varepsilon_t, A_i, \Sigma\}$. Second, I estimate the orthogonality conditions (4) with the GMM using $\left\{\widehat{\varepsilon}_t(\beta), \widehat{A}_i(\beta), \Sigma_T(\beta)\right\}|_{\beta=\beta_T^{VAR}}$ as regressors.

This econometric strategy implies that some variables in the second step are generated regressors from the first step. Therefore, to make statistical inference the asymptotic standard errors calculated from the GMM estimator should be corrected.

To do it, we can compute the asymptotic standard errors of λ in a model that estimates jointly the parameters of the VAR(p) and the SIPC. The problem is described formally in the Appendix. The corrected variance of λ_T is then:

$$V(\lambda_T) = \left[(TV_{na}(\lambda_T))^{-1} - E\frac{\partial g_{1,t}'}{\partial\lambda}\Sigma_{g_1}^{-1}E\frac{\partial g_{1,t}}{\partial\beta'}\cdot \left(E\frac{\partial g_{1,t}'}{\partial\beta}\Sigma_{g_1}^{-1}E\frac{\partial g_{1,t}}{\partial\beta'} + E\frac{\partial g_{2,t}'}{\partial\beta}\Sigma_{g_2}^{-1}E\frac{\partial g_{2,t}}{\partial\beta'}\right)^{-1}\cdot E\frac{\partial g_{1,t}'}{\partial\beta}\Sigma_{g_1}^{-1}E\frac{\partial g_{1,t}}{\partial\lambda}\right]^{-1}/T \quad (5)$$

where $g_{1,t}$ is the vector of orthogonality conditions (4), $g_{2,t}$ is the vector the orthogonality conditions used to estimate the VAR(p) in the first step,[15] β is the vec of the VAR(p) matrices of parameters, and Σ_x is the variance of moments x.

I write the correct variance (5) as function of the not-adjusted one, $V_{na}(\lambda_T)$. It is useful to see that $V(\lambda_T) \geq V_{na}(\lambda_T)$.

Finally, note that if the residuals from the first step estimation are uncorrelated with the ones from the second step, then the two steps estimator is also the most efficient among the GMM estimators of (4).

3.2 VAR estimation

I provide estimates for two main specifications of the $VAR(p)$ model. (i) the baseline, where Z_t includes inflation, output gap and interest rate; (ii) a second one that I named *min RMSE,* where Z_t includes the most relevant variables to forecast inflation and output gap according to Stock and Watson (2003a).[16]

[14] It can be shown that if $\left\{\widehat{\varepsilon}_t, \widehat{A}_i, \Sigma_T\right\}$ are consistent estimators of $\{\varepsilon_t, A_i, \Sigma\}$, then the sample analog

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

converges almost surely to the population moment (4).

[15] The VAR(p) is estimated LS equation by equation.

[16] Stock and Watson (2003a) analyzed the contribution of several variables in forecasting inflation and output gap.

In details, I estimate

$$Z_t = \sum_{j=1}^{p} B_j \cdot Z_{t-j} + \varepsilon_t \tag{6}$$

where

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

and X_t can be either $X_t = i_t$ (the baseline specification), or a $(n-2\times 1)$ vector that includes: short term interest rate (the 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); IMF price index of commodities; real money (real M2 minus small time deposits); unemployment rate; total capacity utilization rate (TCU).

I estimate both the specifications for inflation measured either with CPI, or with the implicit GDP deflator. Output gap is detrended with the HP filter. All the variables are taken in logs except for unemployment, TCU, and interest rates.

The variables have been detrended or taken in first difference when necessary, so all the series used in the VAR(p) are stationary. Also, the VAR(p) model has the minimum number of lags in order for the residuals to be not serially correlated. These two conditions assure that the VAR(p) estimator is consistent.

The sample goes from 1957q1 to 2005q4; The database come from FRED II database of U.S. economy.[17]

3.3 GMM Estimation

To control for the small sample bias problem that affect nonlinear GMM estimators I estimate two alternative specifications of (4). The first one is (4) multiplied by $(1-\lambda)$, the second one is (4) multiplied by $\left(\frac{1-\lambda}{\alpha\lambda}\right)$. They are referred to as (1) and (2) in next tables. The sample goes from 1958q4 to 2005q4 (189 observation).[18]

I provide estimations only for parameter λ, while I calibrate α. In the original model α depends on the intertemporal elasticity of substitution of consumers, on Fisher elasticity of labor, and on the elasticity of demand of single-variety goods. Since I don't use data on neither about consumption or about labor or about firms' markup, then I don't attempt to estimate α.

3.3.1 Results

My first attempt has been to estimate (4) using all the orthogonality conditions, i.e. for $i = 0, ..., l$.

[17] Available at Federal Reserve Bank of St. Louis.

[18] The GMM sample is shorter than the VAR one because I loose 7 observations to obtain the VAR(6) estimates.

The results are not encouraging. We never accept the null hypothesis of overidentifying restrictions in Hansen's J-test, no matter the order of lags l I choose (2, 4, 6, 8, 12), the inflation index I choose (either CPI or GDP deflator), the VAR(p) specification I use to generate the regressors (either the baseline, or the *min RMSE*). Apparently, the model can't match the selected moments all together.

From this first evidence we should argue that there are some sources of mispecification in the SIPC model. However, since the main call of the SIPC was to explain inflation persistence, I check whether the model can do better matching the lagged moments alone. To do this, I discard the first equation in (4), and I focus on the other orthogonality conditions, i.e. equations (4) for $i = 1, ..., l$.

The following table 1 summarizes the results.

Restricted $\alpha = .2$	Specif.	λ_T^{2s}	Adjusted std.err.	Null MR cal.	t-stat (p-val)	Null RE	t-stat (p-val)	J-stat (p-val)
O.C. (4)	$i = 1, ..., 6$							
defl; VAR	(1)	0.35	0.149	0.25	0.71	1	-4.30	2.22
$\{\Delta y_t, \pi_t, i_t\}$					(0.47)*		(0.00)	(0.81)*
	(2)	0.36	0.055	0.25	2.17	1	-11.41	1.95
					(0.03)		(0.00)	(0.85)*
defl; VAR	(1)	0.38	0.106	0.25	1.35	1	-5.70	2.18
minRMSE					(0.17)*		(0.00)	(0.82)*
	(2)	0.41	0.075	0.25	2.15	1	-7.81	2.79
					(0.03)		(0.00)	(0.87)*
cpi; VAR	(1)	0.47	0.061	0.25	3.67	1	-8.44	2.70
$\{\Delta y_t, \pi_t, i_t\}$					(0.00)		(0.00)	(0.74)*
	(2)	0.49	0.057	0.25	4.21	1	-8.85	2.24
					(0.00)		(0.00)	(0.81)*
cpi; VAR	(1)	0.54	0.101	0.25	2.88	1	-4.48	3.38
minRMSE					(0.00)		(0.00)	(0.64)*
	(2)	0.57	0.094	0.25	3.48	1	-4.47	2.74
					(0.00)		(0.00)	(0.73)*

Table 1. 2-step GMM with optimal weighting matrix. U.S. data, sample 1958q4 – 2005q4. Output gap filtered with HP filter. Newey-West HAC standard errors adjusted for stochastic regressors. p-values in parenthesis. J-statistics is Hansen test of overidentifying restrictions (5 d.o.f.).

The estimates are reasonable and quite precise. The model fits very well all the moments according to the J-test. We can never reject the null hypothesis of overidentifying restrictions.

The estimates of the frequency of information updating λ are our main concern. In all specifications λ_T^{2s} is in the range assumed by the theory, i.e. within the $(0, 1]$ interval. More precisely, it ranges between $[0.35, 0.57]$. In column 4 of table 1 I report the p-value for the null hypothesis $\lambda = 0.25$, which is the calibration used by Reis (2004) to match actual inflation persistence with the SIPC model. λ_T^{2s} does not significantly differs from the original MR calibration, at least when we use the GDP deflator as inflation index.

Finally, it seems that the smaller are forecast errors $\{\widehat{\varepsilon}_t\}_{t=1}^T$, the bigger is λ_T^{2s}, as we can see comparing the estimates of the *min RMSE* model against those of the baseline.[19]

Using this restricted set of moments, the results are substantially different from the ones found before. Now the model fit well the data, and the estimates of firms' average information updating comfort MR calibration. We could have expected this result: Reis (2004) showed that $\lambda = 0.25$ is the correct value for the SIPC model to reproduce the persistence of actual U.S. inflation. Therefore, if we force the model to match the moments that measure persistence, then it is likely that the best parameter to do it is indeed the 0.25 proposed by Reis.

It is worth noticing that the estimates in table 1 are in line with the ones of the other empirical papers that estimate the SIPC. For the sake of comparison, I estimate λ using exactly the same information those papers used, which turns out to be the information contained in the first $acf(i)$ of δZ_t for $i = 1, ..., l$.

Using lagged δZ_t as instruments, I derive and estimate the following simulated orthogonality conditions:

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

In this case the estimates of λ_T^{2s} ranges between $[0.30, 0.41]$. This result is close to the one obtained using lagged ε_t as instruments, and resembles much the ones found by Khan and Zhu (2002, 2006), Kiley (2006) Korenok (2005). In conclusion, it seems we can reproduce the same results they found once we use the same information they used, although the estimators are different. Aside, this works as a double check on the correctness of the methodology that I use in this paper.

At this point, the question that naturally follows is whether the SIPC model can match the conditional variance of inflation alone. To do it, I estimate the first orthogonality condition in (4). In order to obtain more precise estimates, that orthogonality condition is multiplied by a vector of instruments x_t, which contains all variables dated $t-1$ and before.[20] Using the additional assumption that the errors $\varepsilon_t's$ are i.i.d. I obtain

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

[19] Recall that the *min RMSE* VAR(p) predicts better Z_t than the baseline specification because it uses more information. Therefore, the residuals $\widehat{\varepsilon}_t$ in the *min RMSE* are smaller than the ones in the baseline VAR(p).

[20] I use a total of 19 instruments, namely: 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.

which is a new set of orthogonality conditions with dimensions 19×1 (the number of instruments used).

(7) is estimated using the same procedure as before. In this case the GMM estimator turns out to be the Non-linear IV estimator, but with smaller variance. To minimize the standard errors of the estimates, the weighting matrix is chosen to be the inverse of the variance of moments. The results are summarized in table 2.

Restricted $\alpha = .2$ O.C. (7)	Specif.	λ_T^{2s}	Adjusted std.err.	Null MR cal.	t-stat (p-val)	Null RE	t-stat (p-val)	J-stat (p-val)
defl; VAR	(1)	0.75	0.089	0.25	5.58	1	-2.76	22.21
$\{\Delta y_t, \pi_t, \mathrm{i}_t\}$					(0.00)		(0.00)	(0.22)*
	(2)	0.86	0.084	0.25	7.24	1	-1.59	15.29
					(0.00)		(0.11)*	(0.64)*
defl; VAR	(1)	0.71	0.099	0.25	4.64	1	-2.90	22.62
minRMSE					(0.00)		(0.00)	(0.205)*
	(2)	0.84	0.103	0.25	5.70	1	-1.55	15.55
					(0.00)		(0.12)*	(0.623)*
cpi; VAR	(1)	0.82	0.073	0.25	7.83	1	-2.43	20.26
$\{\Delta y_t, \pi_t, \mathrm{i}_t\}$					(0.00)		(0.01)	(0.318)*
	(2)	0.85	0.067	0.25	9.03	1	-2.11	13.97
					(0.00)		(0.03)	(0.731)*
cpi; VAR	(1)	0.80	0.083	0.25	6.61	1	-2.39	22.93
minRMSE					(0.00)		(0.01)	(0.197)*
	(2)	0.85	0.079	0.25	7.70	1	-1.79	16.39
					(0.00)		(0.07)*	(0.565)*

Table 2. 2-step GMM with optimal weighting matrix. U.S. data, sample 1958q4 – 2005q4. HP filter for output gap. Newey-West HAC standard errors adjusted for stochastic regressors. p-values in parenthesis. J-statistics is Hansen test of overidentifying restrictions (18 d.o.f.).

Again, the estimates are quite precise and lie in the range assumed by the theory. The model fits quite well the moments according to the J-test. We can never reject the null hypothesis of overidentifying restrictions.[21]

λ_T^{2s} now ranges in the interval $[0.71, 0.86]$, which is significantly higher than the one found in table 2. This value implies that average information duration ranges from 3.5 to around 4 months. As before, I report the p-value for the null hypothesis $\lambda = 0.25$: it is rejected in all the specifications, suggesting that MR

[21] A yellow flag should be lied here. Standard distributions for hypothesis testing with IV estimators are reliable only if the instruments are not weak.

Unfortunately, is still unclear how to check for weak instruments in nonlinear estimators with possibly nonspherical residuals.

calibration should be dismissed if we want the SIPC to match the conditional variance of inflation.

In some specification, however, λ_T^{2s} is close to 1. Since the SIPC with $\lambda = 1$ encompasses RE equilibrium with flexible prices, then I test whether the hypothesis of RE is accepted or rejected by the data. In column 6 of table 2 I report the t-statistics and the p-value for the null hypothesis $\lambda = 1$. The evidences are not overwhelming, the null is rejected at 5% level in most of the specifications, but it is accepted in more than half of them at 1%. We can't say anything about the RE equilibrium, at least from this estimation.

All in all, this second estimation suggests that the SIPC can fit also the conditional variance of inflation, although the model needs an higher frequency λ of firms' information updating to achieve it.

3.3.2 Interpretation of the results and the Hybrid SIPC

Understanding these results is not immediate. Basically, we can say that if we use the SIPC model to match inflation persistence, then it predicts an inflation volatility higher than the one we observe in the data. While if we use the model to match the conditional variance of inflation, then it predicts a lower inflation persistence with respect to the actual one.

A possible explanation of these results is the following. The SIPC has been criticized because it predicts that all producers change the price in every period, while there are robust (across countries and times) evidences that a sizeable fraction of firms change the price only sporadically.

According with this evidence, the SIPC model would be mispecified because it does not take into account that in every period a fraction of producers keep last period price. Would this argument be useful to understand the results above? In a model with both adaptive and inattentive producers the effect of a shock on inflation lasts in time for two reasons. First, since inattentive agents take some periods to get aware of a shock, then a current shock will affect prices in the future. Second, since adaptive producers use lagged prices to set their current price, then a shock that shift inflation today will affect also prices tomorrow. Thus, the covariance between current inflation and lagged shocks in such model depends both on the frequency of information updating and on the size of the fraction of adaptive producers. In particular, that covariance may be high even with a low degree of information stickiness insofar as there is a big fraction of adaptive producers. In this latter case, if we estimated the (mispecified model) SIPC matching the orthogonality conditions (4) with $i = 1, ..., l$ we may find (downward) biased estimates of λ_T.[22]

Following this intuition, it would be useful to derive and test a model with heterogeneous agents, where some of them are inattentive and some others use adaptive pricing rules. Dupor, Kitamura and Tsuruga (2006) goes in this direction. They proposed a model of "dual stickiness" where producers change

[22] It is easy to see that. If the RHS of (4) increases because there are more adaptive producers and inflation is more persistent, helding fixed $y_t, \Delta y_t$ and ε_t, then λ must decrease in order for the equality (4) to hold.

prices only sporadically, plus they absorb the relevant information for pricing in random periods, as in the SIPC.

Basically, what DKT do is to nest together Calvo's sticky price framework with the sticky information of Mankiw and Reis. In DKT model inflation in period t is function of all past periods expectations of a stream of future variables from $t+1$ onwards. It is not immediate to accommodate such inflation dynamics in the framework I use in this paper, and I leave it for future research. Instead, in this paper I do a simpler. I test whether the SIPC model is able to fit the data once we introduce a fraction of pure adaptive firms.

So, I assume that in the economy there are two types of producers. The first type, a fraction φ of adaptive firms, set the price price equal to last period aggregate price level, either adding the inflation over the previous quarter (indexation of prices), or not. The other $(1-\varphi)$ producers are inattentive firms as in the SIPC model.

In this economy the aggregate price index is given by:

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

where

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

Inflation in this "hybrid" model evolves according to:

$$\pi_t = \begin{cases} (1-\varphi)\left[\frac{\alpha\lambda}{1-\lambda} y_t + \lambda \sum_{j=0}^{\infty} (1-\lambda)^j E_{t-j-1}\left(\pi_t + \alpha \Delta y_t\right)\right] \\ \qquad + \varphi \pi_{t-1} & \text{case (a)} \\ (1-\varphi)\left[\frac{\alpha\lambda}{1-\lambda} y_t + \lambda \sum_{j=0}^{\infty} (1-\lambda)^j E_{t-j-1}\left(\pi_t + \alpha \Delta y_t\right)\right] \\ \qquad + \varphi\left(2\pi_{t-1} - \pi_{t-2}\right) & \text{case (b)} \end{cases} \tag{8a}$$

Applying to the models in (8a) the econometric procedure presented in section 3.1, I obtain the following sets of orthogonality conditions:

$$\begin{aligned} E\left[\left(\frac{\alpha\lambda}{1-\lambda} y_t + \alpha \Delta y_t - \frac{\varphi}{1-\varphi}\left(\pi_t - \pi_{t-1}\right)\right)\left(\delta \varepsilon_{t-i}\right)'\right] &= \qquad (9) \\ &= (1-\lambda)^i \delta A_i \Sigma \delta' \\ & \quad \text{case (a)} \\ E\left[\left(\frac{\alpha\lambda}{1-\lambda} y_t + \alpha \Delta y_t - \frac{\varphi}{1-\varphi}\left(\pi_t - 2\pi_{t-1} + \pi_{t-2}\right)\right)\left(\delta \varepsilon_{t-i}\right)'\right] &= \qquad (10) \\ &= (1-\lambda)^i \delta A_i \Sigma \delta' \\ & \quad \text{case (b)} \\ for \quad i &= 0, ..., l \end{aligned}$$

Orthogonality conditions (9) and (10) are estimated by GMM. Analogously with previous estimation, I substitute regressors $\{\varepsilon_t, A_i, \Sigma\}$ with
$\left\{\widehat{\varepsilon}_t(\beta), \widehat{A}_i(\beta), \Sigma_T(\beta)\right\}|_{\beta=\beta_T^{VAR}}$.[23]

Hybrid SIPC $i=0,...,l$ Restr. $\alpha=.2$	Specif.	(a) O.C. (9) λ_T^{2s} (s.e.)	φ_T^{2s} (s.e.)	J-stat (p-val)	(b) O.C. (10) λ_T^{2s} (s.e.)	φ_T^{2s} (s.e.)	J-stat (p-val)
defl; VAR	(1)	0.41	-1.23*	3.56	0.62	-0.66*	3.08
$\{\Delta y_t, \pi_t, i_t\}$		(0.055)	(0.775)	(0.61)*	(0.101)	(0.476)	(0.68)*
	(2)	0.44	-1.33*	3.18	0.67	-0.81*	2.05
		(0.060)	(0.885)	(0.67)*	(0.101)	(0.664)	(0.84)*
defl; VAR	(1)	0.45	-1.08*	1.85	0.64	-0.54*	1.37
minRMSE		(0.054)	(0.813)	(0.86)*	(0.104)	(0.479)	(0.92)*
	(2)	0.46	-1.14*	1.76	0.66	-0.60*	1.03
		(0.056)	(0.884)	(0.88)*	(0.107)	(0.568)	(0.95)*
cpi; VAR	(1)	0.55	-1.37*	8.05	0.75	-0.66*	4.65
$\{\Delta y_t, \pi_t, i_t\}$		(0.065)	(0.986)	(0.15)*	(0.064)	(0.416)	(0.45)*
	(2)	0.64	-2.40*	6.35	0.77	-0.79*	3.39
		(0.080)	(2.628)	(0.27)*	(0.065)	(0.553)	(0.64)*
cpi; VAR	(1)	0.69	-0.83*	6.18	0.78	-0.34*	3.88
minRMSE		(0.087)	(0.671)	(0.28)*	(0.070)	(0.265)	(0.56)*
	(2)	0.78	-0.97*	4.46	0.80	-0.39*	3.21
		(0.101)	(1.09)	(0.48)*	(0.072)	(0.330)	(0.66)*

Table 3. 2-step GMM with optimal weighting matrix. U.S. data, sample 1958q4 – 2005q4. HP filter for output gap. Newey-West HAC standard errors not adjusted for stochastic regressors. * means that the coefficient is not significatively different from zero. J-statistics is Hansen test of overidentifying restrictions (18 d.o.f.).

As we can see in table 3 the good result is that now all the specification are accepted by the data: according to the J-statistics in all specifications the selected moment are "close enough" to zero at parameter estimates. Also, λ_T^{2s} lies in between the estimates found before, implying an average information duration of 6 months, or twice a year, which seems reasonable. On the contrary, the estimates of the fraction of adaptive producers are a reason of concern. The coefficient φ_T^{2s} has the wrong sign in all the specifications, and it is never significantly different from zero.

All in all, the results with the hybrid model don't seem satisfying. The hypothesis of non-maximizers adaptive producers seems not the key issue to improve the sticky information model, and eventually a different explanation should be find.

[23] Notice that the VAR(p) model I used in previous section to obtain $\left\{\widehat{\varepsilon}_t(\beta), \widehat{A}_i(\beta), \Sigma_T(\beta)\right\}$ encompasses both the SIPC model and this hybrid version, since the reduced form AR representations of the two models are observationally equivalent.

3.4 Robustness analysis

I check the robustness of previous results along three dimensions.

1. Whether the results are sensitive to the calibration of α (Table 4)

2. Whether they change using a different filter to get the output gap, i.e. Quadratic Detrend instead of the Hodrick-Prescott (Table 5)

3. Whether a different assumption about the dynamics of Z_t changes the estimates of λ. In particular, I assume that $S_t \equiv (\pi_t + \alpha \Delta y_t) \sim AR(2)$, and I estimate jointly λ and the parameters of the AR(2) model (see the appendix) (Tables 6-7)

The evidences about λ are broadly confirmed. Again we find that the model cannot match all the moments (4) together, but it does a good job in matching a subset of them once we separate the conditional variance from the lagged covariances. The null hypothesis of $\lambda = 0.25$ is always rejected when we match the variance, and accepted in most of the specifications when we match the lagged covariances. Finally, the estimations with $\alpha = 0.1$ point out a a stable *inverse* relationship between α and the estimates of λ.

4 Structural Breaks

During the 1990's the U.S. economy experienced a disinflation accompanied by a fall of inflation persistence.[24] In the SIPC model inflation persistence is inversely related to firms' knowledge about the economy: the more that knowledge is outdated, the more persistent inflation is. Therefore, we might expect that average information duration decreased (λ increased) during the sample, and this would explain the reduction in inflation persistence.

To test the hypothesis of an increase in λ, I perform Andrews's supLM test of structural breaks.[25] 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.

The test is applied to both the estimations reported above, and the results are summarized in table 8.

[24] See Bayoumi and Sgherri (2004) for references.

[25] I choose Andrews's (1993) supLM because it is the most powerful test when timing of (possible) breaks is unknown.

Structural breaks H_0 : *no S.B.*		sup LM	Asymptotic critical values 10%	5%	1%
O.C. (4) with $i = 1, ..., 6$	$\pi_0 = .2$	1.18***	6.80	8.45	11.69
	$\pi_0 = .1$	1.18***	7.63	9.31	12.69
O.C. (7)	$\pi_0 = .2$	2.08***	6.80	8.45	11.69
	$\pi_0 = .1$	1.89***	7.63	9.31	12.69

Table 8. π_0 indicates the percent of each tail cut. SupLM test has non-standard distribution. The asymptotic critical values are given in Andrews (1993). * , ** , *** means significance respectively at 1%, 5%, and 10% level.

According to Andrews's test there is an overwhelming evidence that no structural break to λ occurred during the sample. However, there is one reason of concern with this result. During late 1970's inflation volatility increased sharply because of the oil shock, which was an exogenous event with respect to this analysis, but it could possibly biases the results of Andrews test. To see it, in Figure (2) I plot the residuals from the estimation of (4).[26]

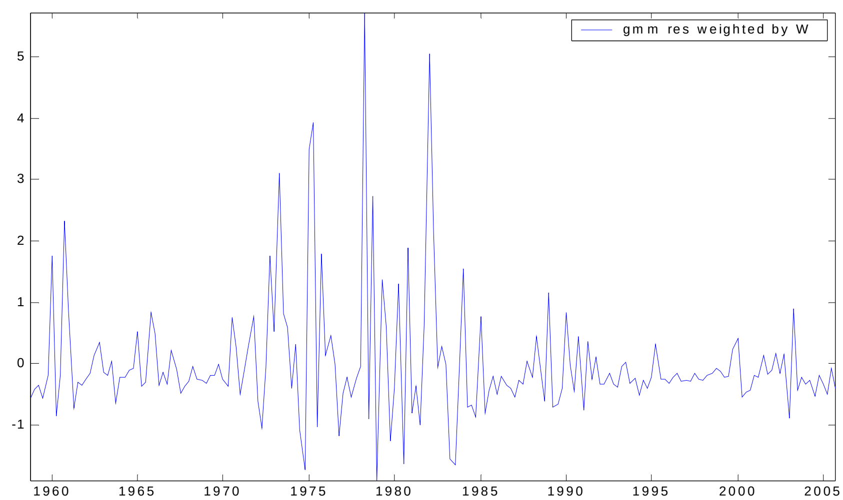

Figure 4.

Therefore, I perform a second test controlling for the effect of the oil shock in the 1970's. Under the null hypothesis that the same model holds throughout the sample, I test whether λ is equal in two subsamples: one that goes from 1959q1

[26] The figure refers to the estimation of (4) with i=0. The figure for the estimation of (4) with i=1,...,l is similar and it is not reported here.

to 1970q1, and the second from 1990q4 to 2005q4. I compare the coefficients using LM and Wald tests. Results are in table 9.

Structural breaks $H_0 : no\ S.B.$	λ_{60} (s.e.)	λ_{90} (s.e.)	$Wald$ (p-val)	LM (p-val)
O.C. (4) with $i = 1, ..., 6$	0.44 (0.069)	0.74 (0.140)	3.83 (0.05)**	5.25 (0.03)*
O.C. (7)	0.61 (0.049)	0.95 (0.025)	44.51 (0.00)	1871.1 (0.00)

Table 9. Wald and LM tests both have standard χ^2 distribution with 1 d.o.f. * , ** , *** means significance respectively at 1%, 5%, and 10% level.

The null hypothesis $\lambda_{60}^{Gmm} = \lambda_{90}^{Gmm}$ is now rejected almost always at 5% level. According to this test, it is likely that a structural break have occurred between the first and the last years of the sample. Now, it make sense that in recent years firms have had better knowledge of the economy and the markets they operate in than 40 years ago. In the past information is likely to have been stickier: less media to channel macroeconomic news, less accurate forecasts about the markets, less experienced authorities, less data gathering, etc. It is not surprising that firms acquired the relevant information slower, therefore taking longer time to react to new events.

5 Conclusions

In this paper I find that SIPC is not a valid model to explain U.S. post-war inflation dynamics. The main reason is that the model cannot explain at the same time inflation persistence and inflation variance. In particular, if we use the model to match inflation persistence, then it predicts an inflation volatility higher than what we observe in the data; while if we use the model to match the conditional variance of inflation, then it predicts a lower inflation persistence with respect to the actual one.

However, once we estimate the model matching only the covariances between current inflation and lagged shocks, the estimates of the frequency of information updating are in accordance with that of the other papers that estimate the SIPC using limited information estimators. I show that this is due to the fact that we are exploiting the same information about the inflation process. In this case I find $\lambda_T^{2s} \in [0.35, 0.57]$. This value implies an average information duration from 6 to 9 months.

Differently, once we use the model to match the conditional variance of inflation I find $\lambda_T^{2s} \in [0.71, 0.86]$. This value implies an average information duration from 3.5 to 4 months, which is just slightly higher than the average information duration in the neoclassical model with RE and flexible prices.

Also, using one set of moments at the time there is a robust evidence that firms' average information duration was significantly higher in the first years of the sample (1960's) than in the last ones (1990's). This finding suggests that sticky information might have been an important source of inflation persistence in past times, while today is not anymore.

All in all, my analysis suggests that sticky information theory cannot be the main explanation of inflation persistence, at least as it is modeled in the SIPC.

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] 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

[8] Coibion, Olivier (2006): "Testing the Sticky Information Phillips Curve", manuscript University of Michigan

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

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

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

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

[13] 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

[14] ——————————— (2006): "Estimates of the Sticky-Information Phillips Curve for the United States", Journal of Money, Credit, and Banking vol. 38 n^o 1, pp. 195-207

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

[16] Laforte, Jean-Philippe (2005): "Pricing Models: A Bayesian DSGE approach for the US", manuscript Federal Reserve Board

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

[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] Mankiw Gregory N. and Reis Ricardo (2006): "Sticky Information in General Equilibrium", manuscript Princeton University

[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 (2003a): "Forecasting Output and Inflation: The Role of Asset Prices", Journal of Economic Literature vol. XLI, pp. 788-829

[26] ——————————————- (2003b): "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 Sticky information and inflation persistence

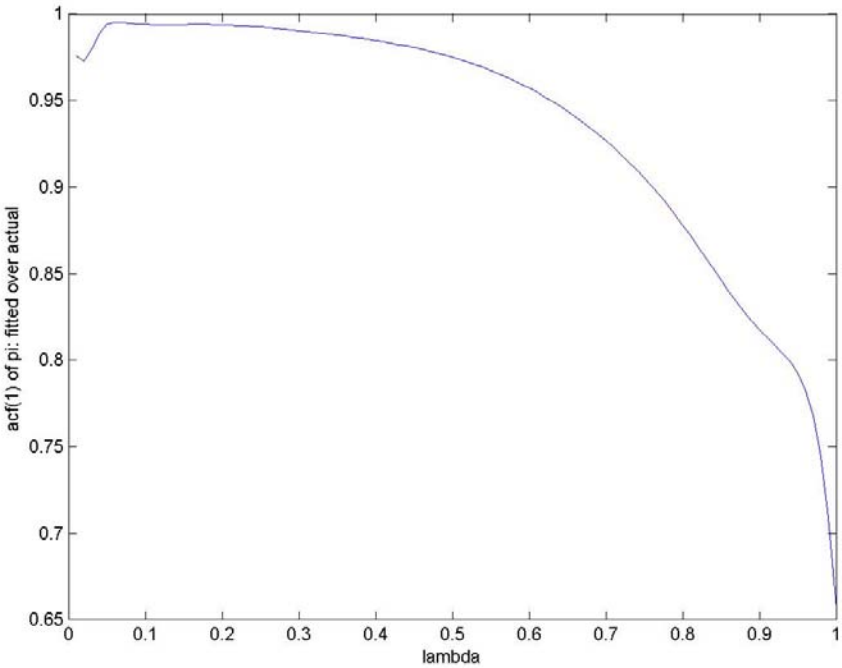

Figure 1.

B Proof of Lemma 1

In order to write the Sticky Information Phillips Curve,

$$\pi_t = \frac{\alpha\lambda}{1-\lambda} y_t + \lambda \sum_{j=0}^{\infty} (1-\lambda)^j E_{t-1-j} [\pi_t + \alpha \Delta y_t] \tag{11}$$

as function of the exogenous shocks, first we define the j-periods-ahead forecast error as:

$$\varepsilon^F_{t|t-j} = Z_t - E[Z_t \mid \Omega_{t-j}] \tag{12}$$

Then, using (12) to substitute out the expectations in (11) we obtain:

$$\pi_t = \frac{\alpha\lambda}{1-\lambda} y_t + \lambda \sum_{j=0}^{\infty} (1-\lambda)^j \delta \left(Z_t - \varepsilon^F_{t|t-j-1} \right) \tag{13}$$

where δ is a $(1 \times n)$ row vector that picks $(\pi_t + \alpha \Delta y_t)$ within Z_t. Equation (13) can be written as:

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

Let's focus now on the RHS of (14). Using the Wold decomposition of Z_t,

$$Z_t = c + \sum_{i=0}^{\infty} A_i \varepsilon_{t-i} \tag{15}$$

we have:

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

Thus, using (15) to substitute out $\varepsilon^F_{t|t-j-1}$ in the RHS of (14) we find:

$$\lambda \sum_{j=0}^{\infty} (1-\lambda)^j \delta \varepsilon^F_{t|t-1-j} = \lambda \sum_{j=0}^{\infty} (1-\lambda)^j \delta \sum_{i=0}^{j} A_i \varepsilon_{t-i}$$

$$= \left(\delta\varepsilon_t + (1-\lambda)\delta\varepsilon_t + (1-\lambda)^2 \delta\varepsilon_t + \ldots \right) + \left((1-\lambda)\delta A_1 \varepsilon_{t-1} + (1-\lambda)^2 \delta A_1 \varepsilon_{t-1} + \ldots \right) + \ldots$$

$$= \frac{\lambda}{1-(1-\lambda)} \sum_{i=0}^{\infty} (1-\lambda)^i \delta A_i \varepsilon_{t-i} \tag{17}$$

Finally, plugging (17) into (14) 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 Adjusted $V\left(\lambda_T^{2s}\right)$

The problem is defined as follows: let

$$E\left[g_1\left(\lambda,\beta,y_t\right)\right]=0 \tag{18}$$

be the set of orthogonality conditions (4), where β is the vec of the matrices of parameters in (6).

λ_T^{2s} in section 3.3 is obtained estimating the sample analog

$$\frac{1}{T}\sum_{t=1}^{T}\left[g_1\left(\lambda,\beta_T^{Var},y_t\right)\right]=0$$

where β_T^{Var} are estimated coefficients of a pth-order vector autoregression model (henceforth VAR(p)) with errors $\varepsilon_t\sim i.i.d.\ N\left(0,\Sigma\right)$. This VAR(p) model has m endogenous variables and it is estimated LS equation by equation. The vector β_T^{Var} has $(m\left(mp+1\right)\times 1)$ elements.

Also, let

$$E\left[\underset{km\times 1}{g_2\left(\beta,y_t\right)}\right]=0 \tag{19}$$

be the orthogonality conditions that we use to estimate β_T^{Var} in the first step, where $k=mp+1$.

Now, to estimate jointly $\{\lambda,\beta\}$ we could stack (18) and (19) in a $(km+n)\times 1$ vector of moments and estimate it by GMM, i.e.

$$E\begin{bmatrix}\underset{n\times 1}{g_1\left(\lambda,\beta,y_t\right)}\\ \underset{km\times 1}{g_2\left(\beta,y_t\right)}\end{bmatrix}=0 \tag{20}$$

In this model there are no stochastic regressors, and the (correct) VCV matrix of coefficients if optimal weighting matrix is used is

$$V\begin{pmatrix}\lambda_T\\ \beta_T\end{pmatrix}=\left[\underbrace{TE\left(\frac{\partial\begin{pmatrix}g_{1,t}\\ g_{2,t}\end{pmatrix}}{\partial\begin{pmatrix}\lambda & \beta'\end{pmatrix}}\right)'}_{\equiv G}\cdot\right.$$

$$\left.\underbrace{\left(E\begin{pmatrix}g_{1,t}\\ g_{2,t}\end{pmatrix}\begin{pmatrix}g'_{1,t} & g'_{2,t}\end{pmatrix}\right)}_{\equiv\Omega}^{-1}E\left(\frac{\partial\begin{pmatrix}g_{1,t}\\ g_{2,t}\end{pmatrix}}{\partial\begin{pmatrix}\lambda & \beta'\end{pmatrix}}\right)\right]^{-1}$$

where G is a $(km+n)\times(km+1)$ matrix, and Ω is the $(km+n)\times(km+n)$ VCV matrix of moments.

By construction $\frac{\partial g_{2,t}}{\partial \lambda} = 0$, so matrix G can be written as

$$G = E \begin{pmatrix} \frac{\partial g_{1,t}}{\partial \lambda} & \frac{\partial g_{1,t}}{\partial \beta'} \\ 0 & \frac{\partial g_{2,t}}{\partial \beta'} \end{pmatrix}$$

In addition, if we assume no covariance between SIPC and VAR residuals,[27] then the inverse of the variance of moments is

$$\Omega^{-1} = \begin{pmatrix} \Sigma_{g_1}^{-1} & 0 \\ 0 & \Sigma_{g_2}^{-1} \end{pmatrix} \tag{21}$$

Using these two facts (**??**) is:

$$V\begin{pmatrix} \lambda_T \\ \beta_T \end{pmatrix} = \left[E \begin{pmatrix} \frac{\partial g_1'}{\partial \lambda} & 0' \\ \frac{\partial g_1'}{\partial \beta} & \frac{\partial g_2'}{\partial \beta} \end{pmatrix} \begin{pmatrix} \Sigma_{Gmm}^{-1} & 0 \\ 0 & \Sigma_{Var}^{-1} \end{pmatrix} E \begin{pmatrix} \frac{\partial g_1}{\partial \lambda} & \frac{\partial g_1}{\partial \beta'} \\ 0 & \frac{\partial g_2}{\partial \beta'} \end{pmatrix} \right]^{-1} /T \tag{22}$$

or, after some algebra manipulation,

$$V\begin{pmatrix} \lambda_T \\ \beta_T \end{pmatrix} = \begin{pmatrix} E\frac{\partial g_{1,t}'}{\partial \lambda}\Sigma_{g_1}^{-1}E\frac{\partial g_{1,t}}{\partial \lambda} & E\frac{\partial g_{1,t}'}{\partial \lambda}\Sigma_{g_1}^{-1}E\frac{\partial g_{1,t}}{\partial \beta'} \\ E\frac{\partial g_{1,t}'}{\partial \beta}\Sigma_{g_1}^{-1}E\frac{\partial g_{1,t}}{\partial \lambda} & E\frac{\partial g_{1,t}'}{\partial \beta}\Sigma_{g_1}^{-1}E\frac{\partial g_{1,t}}{\partial \beta'} + E\frac{\partial g_{2,t}'}{\partial \beta}\Sigma_{g_2}^{-1}E\frac{\partial g_{2,t}}{\partial \beta'} \end{pmatrix}^{-1} /T \tag{23}$$

Now, the not-adjusted variance of the two steps estimator λ_T^{2s} obtained from the estimation of the orthogonality conditions (18) alone is:

$$V_{na}(\lambda_T^{2s}) = \left(TE\frac{\partial g_{1,t}'}{\partial \lambda}\Sigma_{g_1}^{-1}E\frac{\partial g_{1,t}}{\partial \lambda} \right)^{-1} \tag{24}$$

So, using the definition (24) and the formula for the inverse of partitioned matrices, the variance of λ_T in (23) can be written as

$$V(\lambda_T) = \left(\left(TV_{na}(\lambda_T^{2s})\right)^{-1} - E\frac{\partial g_{1,t}'}{\partial \lambda}\Sigma_{g_1}^{-1}E\frac{\partial g_{1,t}}{\partial \beta'} \left(E\frac{\partial g_{1,t}'}{\partial \beta}\Sigma_{g_1}^{-1}E\frac{\partial g_{1,t}}{\partial \beta'} + E\frac{\partial g_{2,t}'}{\partial \beta}\Sigma_{g_2}^{-1}E\frac{\partial g_{2,t}}{\partial \beta'} \right)^{-1} E\frac{\partial g_{1,t}'}{\partial \beta}\Sigma_{g_1}^{-1}E\frac{\partial g_{1,t}}{\partial \lambda} \right)^{-1} /T \tag{25}$$

Thus, (25) is the variance of the estimator of λ in this model, and it turns out to be also the variance of λ_T^{2s} once we adjust for the stochastic regressors.

Actually, we don't need to estimate (20) to get (25), since we can compute it with the information we have from the two steps estimator. In particular, $\frac{\partial g_{1,t}}{\partial \lambda}$ and $\Sigma_{g_1}^{-1}$ are respectively the Jacobian and the weighting matrix of the second step. $\left(E\frac{\partial g_{2,t}'}{\partial \beta}\Sigma_{g_2}^{-1}E\frac{\partial g_{2,t}}{\partial \beta'}\right)$ is the VCV matrix of the VAR(p) parameters in the

[27] It can be shown that this is indeed the case when the errors ε_t are normally distributed.

first step, and $\frac{\partial g_{1,t}}{\partial \beta'}$ is the vector of derivatives of (4) with respect to β evaluated at $\beta_T^{Var}, \lambda_T^{2s}$.

It is worth noticing that (25) is the correct variance of λ_T only if the covariance between $g_{1,t}$ and $g_{2,t}$ is zero. Otherwise, it would be optimal to estimate jointly λ, β because we could exploit the information in $g_{1,t}$ to pin down β_T. But this means a non linear optimization over a large set of parameters, while in this paper I do simpler: I estimate β by OLS, then I estimate λ by GMM, using $\Sigma_{g_1}^{-1}$ as weighting matrix. This is akin to estimate (20) with the GMM using

$$W = \begin{pmatrix} \Sigma_{g_1}^{-1} & O \\ O' & I_{km} \end{pmatrix} \tag{26}$$

as weighting matrix.

Now, if $E\left[g_{2,t} g_{1,t}\right] \neq 0$ then Ω^{-1} in (21) is not diagonal, and (26) is not the optimal weighting matrix to estimate (20). In this case, the correct variance of λ_T is the upper left cell of:

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

and the two-steps estimator is no longer the most efficient estimator of λ_T among GMM estimators of (20).

D Robustness Analysis: Tables

- Calibration of α.

Restricted $\alpha = .1$	Specif.	λ_T^{2s}	Adjusted std.err.	Null MR cal.	t-stat (p-val)	Null RE	t-stat (p-val)	J-stat (p-val)
O.C. (4)	$i = 1, ..., 6$							
defl; VAR	(1)	0.51	0.071	0.25	3.74	1	-5.18	1.91
$\{\Delta y_t, \pi_t, i_t\}$					(0.00)		(0.00)	(0.86)*
	(2)	0.53	0.073	0.25	3.82	1	-6.36	1.62
					(0.00)		(0.00)	(0.89)*
defl; VAR	(1)	0.54	0.099	0.25	2.93	1	-4.64	1.76
minRMSE					(0.00)		(0.00)	(0.88)*
	(2)	0.55	0.104	0.25	2.90	1	-4.28	1.54
					(0.00)		(0.00)	(0.90)*
O.C. (7)								
defl; VAR	(1)	0.93	0.047	0.25	14.42	1	-1.38	17.47
$\{\Delta y_t, \pi_t, i_t\}$					(0.00)		(0.16)*	(0.49)*
	(2)	0.94	0.047	0.25	14.54	1	-1.12	13.39
					(0.00)		(0.26)*	(0.76)*
defl; VAR	(1)	0.92	0.06	0.25	11.14	1	-1.23	20.61
minRMSE					(0.00)		(0.21)*	(0.29)*
	(2)	0.94	0.058	0.25	11.75	1	-1.02	14.11
					(0.00)		(0.30)*	(0.72)*

Table 3. Calibration of α. 2-step GMM estimator with optimal weighting matrix. Orthogonality conditions (4). U.S. data, sample 1958q4 – 2005q4. HP filter for output gap. Newey-West HAC standard errors adjusted for stochastic regressors. p-values in parenthesis. J-statistics is Hansen test of overidentifying restrictions.

- Quadratic detrend (QD) filter.

Restricted $\alpha = .2$	Specif.	λ_T^{2s}	Adjusted std.err.	Null MR cal.	t-stat (p-val)	Null RE	t-stat (p-val)	J-stat (p-val)
O.C. (4)	$i = 1, ..., 6$							
defl; VAR	(1)	0.29	0.055	0.25	0.83	1	-12.61	1.35
$\{\Delta y_t, \pi_t, i_t\}$					(0.40)*		(0.00)	(0.92)*
	(2)	0.30	0.040	0.25	1.31	1	-17.37	1.17
					(0.18)*		(0.00)	(0.94)*
defl; VAR	(1)	0.33	0.099	0.25	0.84	1	10.73	1.02
minRMSE					(0.40)*		(0.00)	(0.96)*
	(2)	0.34	0.060	0.25	1.56	1	10.73	1.29
					(0.11)*		(0.00)	(0.93)*
O.C. (7)								
defl; VAR	(1)	0.64	0.081	0.25	4.82	1	-4.38	29.99
$\{\Delta y_t, \pi_t, i_t\}$					(0.00)		(0.00)	(0.03)
	(2)	0.78	0.090	0.25	5.89	1	-2.37	10.88
					(0.00)		(0.00)	(0.89)*
defl; VAR	(1)	0.58	0.11	0.25	3.01	1	-3.77	35.44
minRMSE					(0.00)		(0.00)	(0.00)
	(2)	0.90	0.15	0.25	4.22	1	-0.58	11.87
					(0.00)		(0.56)*	(0.85)*

Table 4. QD 2-step GMM estimator with optimal weighting matrix. Orthogonality conditions (4). U.S. data, sample 1958q4 – 2005q4. Quadratic Detrend filter for output gap. Newey-West HAC standard errors adjusted for stochastic regressors. p-values in parenthesis. J-statistics is Hansen test of overidentifying restrictions.

- Univariate process for the exogenous shocks, i.e. $(\pi_t + \alpha \Delta y_t) \sim AR(2)$

Let's define $S_t \equiv (\pi_t + \alpha \Delta y_t)$. I assume here that demeaned inflation and output gap follow an univariate second order autoregressive process, i.e.

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

This assumption simplifies the model at issue, and we can estimate jointly the parameters of the $AR(2)$ process and the firms' frequency of information updating parameter λ.

To do that, I estimated jointly the following orthogonality conditions:

$$\begin{aligned} E\left[\begin{pmatrix} S_{t-1} \\ S_{t-2} \end{pmatrix} \cdot (S_t - \phi_1 S_{t-1} - \phi_2 S_{t-2})\right] &= 0 \\ E\left[(S_t - \phi_1 S_{t-1} - \phi_2 S_{t-2})^2 - \sigma_\varepsilon^2\right] &= 0 \\ E\left[\left(\frac{\alpha\lambda}{1-\lambda} y_t + \alpha \Delta y_t\right) \cdot \varepsilon_{t-i} - (1-\lambda)^i A_i \sigma_\varepsilon^2\right] &= 0 \end{aligned} \tag{28}$$

In Table 5 I estimate all the moments (28) together. The results are similar to that found before, when I estimate the orthogonality conditions (4) all together. The moments are rejected by the data. In other words this model cannot fit inflation variance and perisistence together.

Restricted		λ_T^{Gmm}	ϕ_1	ϕ_2	σ_ε^2	H_0: $\lambda = .25$	J-stat
$\alpha = .2$	Specif.	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(p-val)	(p-val)
O.C. (4)	$i = 0, ..., 6$						
GDP deflator	(1)	0.90	0.67	0.28	1.2e-5	7.00	11.24
$\pi_t + \alpha\Delta y_t$ ~AR(2)		(.093)	(.062)	(.065)	(0.2e-5)	(0.00)	(0.08)*
	(2)	0.24	0.69	0.27	0.2e-5	-0.07	23.88
		(.030)	(.063)	(.067)	(0.1e-5)	(0.94)*	(0.00)
CPI	(1)	0.70	0.62	0.32	1.3e-5	5.84	17.72
$\pi_t + \alpha\Delta y_t$ ~AR(2)		(.077)	(.047)	(.044)	(0.3e-5)	(0.00)	(0.00)
	(2)	0.83	0.60	0.32	2.0e-5	9.97	9.04
		(.058)	(.046)	(.046)	(0.3e-5)	(0.00)	(0.17)*

Table 5. Forecast technology. 1-step GMM with optimal weighting matrix. U.S. data, sample 1958q4 – 2005q4. HP filter for output gap. Newey-West HAC standard errors (no stochastic regressors). p-values in parenthesis. J-statistics is Hansen test of overidentifying restrictions.

However, following the same strategy as before, I discard first the conditional variance and I estimate the other moments, and then I discard the moments related with the lagged covariances and I estimate the contemporaneous covariance alone. Results are in Table 6.

Restricted $\alpha = .2$	Specif.	λ_T^{Gmm} (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)
O.C. (4)	$i = 1, ..., 6$						
GDP deflator	(1)	0.51	0.61	0.33	1.4e-5	3.92	0.97
$\pi_t + \alpha\Delta y_t$ ~AR(2)		(.067)	(.061)	(.064)	(0.2e-5)	(0.00)	(0.96)*
	(2)	0.52	0.62	0.32	1.4e-5	4.02	0.83
		(.068)	(.060)	(.063)	(0.2e-5)	(0.00)	(0.97)*
CPI	(1)	0.54	0.66	0.28	1.9e-5	6.74	2.27
$\pi_t + \alpha\Delta y_t$ ~AR(2)		(.043)	(.051)	(.048)	(0.3e-5)	(0.00)	(0.80)*
	(2)	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. (7)							
GDP deflator	(1)	0.91	0.61	0.34	1.4e-5	7.66	exacly
$\pi_t + \alpha\Delta y_t$ ~AR(2)		(.086)	(.066)	(.074)	(0.2e-5)	(0.00)	identif.
	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)*
CPI	(1)	0.81	0.66	0.28	2.1e-5	9.59	exacly
$\pi_t + \alpha\Delta y_t$ ~AR(2)		(.058)	(.060)	(.056)	(0.4e-5)	(0.00)	identif.
	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)*

Table 6. Forecast technology. 1-step GMM with optimal weighting matrix. U.S. data, sample 1958q4 – 2005q4. HP filter for output gap. Newey-West HAC standard errors (no stochastic regressors). p-values in parenthesis. J-statistics is Hansen test of overidentifying restrictions.