CARF Working Paper

CARF-F-516

Identifying the Source of Information Rigidities in the Expectations Formation Process

Mototsugu Shintani
University of Tokyo

Kozo Ueda
Waseda University
and Centre for Applied MacroeconomicAnalysis (CAMA)

June 3, 2021

CARF is presently supported by Nomura Holdings, Inc., Sumitomo Mitsui Banking Corporation, The Dai-ichi Life Insurance Company, Limited, The Norinchukin Bank, MUFG Bank, Ltd. and Ernst & Young ShinNihon LLC. This financial support enables us to issue CARF Working Papers.

Identifying the Source of Information Rigidities in the Expectations Formation Process

Mototsugu Shintani* Kozo Ueda†

June 3, 2021

Abstract

Coibion and Gorodnichenko (2015) provide a useful framework to test the null hypothesis of full-information rational expectations against two popular classes of information rigidities, sticky information (SI) and noisy information (NI). However, the observational equivalence of SI and NI in average forecast errors gives no power in the test for one against the other. We identify the source of information rigidities by estimating the equations for the average forecast errors and variance of forecasts. The results show the importance of both SI and NI, but favor a type of NI in which agents quickly learn the underlying state.

Keywords: imperfect information; heterogeneity; sticky information; noisy information; observational equivalence

JEL classification: C53, D83, D84, E13, E31, E37

---

*University of Tokyo (Email: shintani@e.u-tokyo.ac.jp)

†Waseda University and Centre for Applied MacroeconomicAnalysis (CAMA) (Email: kozo.ueda@waseda.jp). The authors would like to thank Toshihiko Mukoyama and seminar participants at the University of Tokyo and gratefully acknowledge the financial support of Grant-in-aid for Scientific Research (19H01491 and 20H01482).

1 Introduction

Coibion and Gorodnichenko (2015) introduce a useful regression framework to test the null hypothesis of the full-information rational expectations against the alternative hypothesis of rational expectations in the presence of information rigidities. In particular, their proposed test, which is based on a regression of average forecast errors on the average forecast revision, has power against two popular classes of information rigidities frequently employed in the macroeconomic literature: sticky information (SI) and noisy information (NI). Once deviation from the full-information model is confirmed by data using Coibion and Gorodnichenko's regression, it is natural to examine which one of the two classes of information rigidities is more appropriate in describing the actual expectation formation process in the next step. However, the observational equivalence of SI and NI in their regression makes it impossible to construct a testing procedure to distinguish between the two. In other words, a test for the null hypothesis of NI has no power against an alternative of SI, and vice versa.

In this paper, we discuss the identification issue of the two sources of information rigidities in Coibion and Gorodnichenko's regression framework. To clarify the issue, we first construct a simple hybrid model of SI and NI and show the implications of the model for the cross-sectional average of forecast errors. Second, we show the implications of the hybrid model for the cross-sectional variance of forecasts. We then shut down one of the two sources of information rigidity to consider the conditions for identifying SI from NI and/or NI from SI.

The three main outcomes from our analytical exercise are as follows. First, the identification of the source of information rigidity based solely on the cross-sectional average of forecast errors crucially depends on the speed of learning in NI models. Specifically, if an underlying state becomes common knowledge at the end of each period, as in the case of Lucas (1972), Angelotos and La'O (2009), and Crucini, Shintani, and Tsuruga (2015), among others, additional terms appearing in Coibion and Gorodnichenko's regression enable the identification of NI from SI. This result is in contrast with the observational equivalence of SI and NI in their regression when realizations are never revealed to agents in the NI model, as in the case of Woodford (2003) and Coibion and Gorodnichenko (2012, 2015), among others. Second, as appropriately pointed out by Coibion and Gorodnichenko (2012), the dependence of disagreements among agents on aggregate shocks is a useful feature in distinguishing NI from SI. We further derive a simple analytical form for the cross-sectional variance that is useful in identifying SI from NI, irrespective of the specification of NI. Accordingly, testing the null hypothesis of NI based on cross-sectional variance has a power against an alternative of SI, and vice versa.[1] Third, we point out that the joint estimation of the two equations, one for average forecast errors and the other for cross-sectional variance, is helpful, not only from the perspective of identification but also from the efficiency in estimating the structural parameters.

Basing our analysis on these considerations, we revisit the empirical findings of Coibion and Gorodnichenko (2015) by using the same dataset on inflation forecasts from the U.S. Survey of Professional Forecasters (SPF).[2] The results of our empirical analyses are summarized as follows. First, from the single-equation estimation for average forecast errors, we find that the null hypothesis of pure SI is not rejected against the alternative of pure Lucas (1972)-type NI. Second, from the single-equation estimation for cross-sectional variance, we find that the null hypothesis of pure NI is rejected against the alternative of pure SI, while the null hypothesis of pure SI is also rejected against the alternative of pure NI. Third, from the joint estimation of the two equations for pure models of information rigidities, the nonnested test suggests that the null hypothesis of the pure SI model is not rejected against both types of pure NI models, while the

[1]Andrade and Le Bihan (2013) and Hur and Kim (2016) also use cross-sectional variance to detect NI.

[2]Most empirical studies focus on either of the two information rigidities. For example, pure SI is studied by Mankiw, Reis, and Wolfers (2004), Branch (2007), Crucini, Shintani, and Tsuruga (2010), and Armantier et al. (2016), while pure NI is studied by Crucini, Shintani, and Tsuruga (2015). An exception is Andrade and Le Bihan (2013), who consider a hybrid model. However, they do not derive simple analytical forms as we do.

null hypotheses of the two types of pure NI models are significantly rejected against the alternative of other pure models of information rigidities. Fourth, from the joint estimation of the two equations for hybrid models of information rigidities, we find a nonnegligible degree of information stickiness. At the same time, while the information noise is relatively small, the formulation of Lucas (1972) better fits the data than that of Woodford (2003) for the NI part of the hybrid model.

The remainder of this study is structured as follows. We discuss the main implications of our model of information rigidities on the average forecast errors and cross-sectional variance of forecasts in Sections 2 and 3, respectively. The joint estimation of the average and variance equations for pure models of information rigidities is conducted in Section 4, followed by the joint estimation for hybrid models of information rigidities in Section 5. Section 6 concludes our discussion.

2 Cross-sectional Average Forecast Errors

2.1 Models for Cross-sectional Average Forecast Errors

Following the analysis of Coibion and Gorodnichenko (2015), we focus on the two classes of information rigidities, SI and NI. For the purpose of clarifying the identification issue, in what follows, we introduce a hybrid model of SI and NI. See Online Appendix A for the detailed derivation of the results.

Sticky Information and Woodford-type Noisy Information

We let the inflation rate π_t follow a stationary AR(1) process that is given by $\pi_t = \rho\pi_{t-1} + \nu_t$, where $|\rho| < 1$ and ν_t represents an i.i.d. normal shock with mean zero. As for NI, we follow Woodford (2003) and assume that agent i cannot observe the current and past values of π_t directly in period t. She can instead receive her individual signal π_{it} in period t, where $\pi_{it} = \pi_t + \omega_{it}$ and ω_{it} is the mean-zero normal noise, which is i.i.d. across time t and agent i with variance σ^2_ω. As for SI, we assume that agent i can update her information set and revise her forecasts with probability $1-\lambda$, where $0 \leq \lambda < 1$. Agent i who revises her expectation in period t can incorporate current and past individual signals π_{it-j} for all $j \geq 0$. The cross-sectional average of h-period ahead forecasts in period t is then given by

$$F_t\pi_{t+h} = (1-\lambda)\sum_{j=0}^{\infty} \lambda^j \mathcal{F}_{t-j}\pi_{t+h},$$

where $\mathcal{F}_t\pi_{t+h}$ denotes the cross-sectional average of $\mathcal{F}_{it}\pi_{t+h}$, namely, the forecast of agent i who revises her forecast in period t. By the recursive substitution of the AR(1) structure, we have

$$\begin{aligned}\mathcal{F}_{it}\pi_{t+h} &= \rho^h \mathcal{F}_{it}\pi_t = \rho^h G\pi_{it} + \rho^h(1-G)\mathcal{F}_{it-1}\pi_t \\ &= \rho^h G\pi_{it} + (1-G)\mathcal{F}_{it-1}\pi_{t+h}\end{aligned}$$

where G denotes the Kalman gain ($0 < G \leq 1$), which takes a value one when the noise is absent in the signal ($\omega_{it} = 0$). The average forecast error among the agents who revise their forecasts is given by

$$\pi_{t+h} - \mathcal{F}_t\pi_{t+h} = (1-G)(\pi_{t+h} - \mathcal{F}_{t-1}\pi_{t+h}) + G\nu_{t+h,t},$$

where $\nu_{t+h,t}$ is the weighted sum of the future shocks in the AR(1) process, from ν_{t+1} to ν_{t+h}. Here $\nu_{t+h,t}$ is uncorrelated with all the information dated t or earlier. Most importantly, the average *ex post* forecast error (FE) in the hybrid model is given by

$$\pi_{t+h} - F_t\pi_{t+h} = \frac{1-(1-\lambda)G}{(1-\lambda)G}(F_t\pi_{t+h} - F_{t-1}\pi_{t+h}) - \frac{(1-G)\lambda}{(1-\lambda)G}(F_{t-1}\pi_{t+h} - F_{t-2}\pi_{t+h}) + \nu_{t+h,t}, \tag{1}$$

which depends both on the current and lagged *ex ante* forecast revisions (FRs). Clearly, equation (1) nests the cases of pure SI ($\lambda > 0$ and $1-G = 0$) and pure NI ($\lambda = 0$ and $1-G > 0$). In particular, the equation reduces to equation (5) of Coibion and

Gorodnichenko (2015) when $1-G=0$, while it reduced to their equation (9) when $\lambda=0$. This relationship suggests that pure SI and NI models are observationally equivalent. If the information structure is one of the two models, a test for the null hypothesis of pure NI (or SI) has no power against an alternative of pure SI (or NI) in the framework of Coibion and Gorodnichenko's regression.[3]

It should also be noted that the second term on the right-hand side of the equation does not show up in Coibion and Gorodnichenko's (2015) regression. However, when information is both sticky and noisy, the coefficient for the second term is non-zero, so that the average FE in period t depends on the FR in period $t-1$ (the second term on the right-hand side of the equation) as well as that in period t (the first term). Thus, in principle, we can obtain the values of $1-G$ and λ separately from the two coefficients. However, if one of the two parameters on the degree of information rigidity is small, namely, either $1-G$ or λ is close to zero, the second coefficient will be small and the model is only weakly identified. This situation poses an econometric challenge for the identification between the two pure models of information rigidities, namely, the model with only SI ($1-G=0$) and the model with only NI ($\lambda=0$). As we will discuss in Section 3, the identification of Woodford-type NI calls for the use of cross-sectional variance or joint estimation.[4]

[3]Recently, Fuhrer (2018), Angeletos, Huo and Sastry (2020), Bordalo et al. (2020), and Broer and Kohlhas (2020), have investigated Coibion and Gorodnichenko's regression at the individual level by regressing individual FE on individual FR, rather than regressing average FE on average FR. Kohlhas and Walther (2021), on the other hand, regress individual FE on average FR. However, since a non-zero coefficient on FR implies the irrationality of forecasters irrespective of information rigidities, running the regression at the individual level does not solve the issue of identifying the source of information rigidities.

[4]An equation similar to equation (1) is also estimated by Coibion and Gorodnichenko (2015) in the context of forecast smoothing. They estimate the equation using two lags of log changes in oil prices as instruments. In Column (3) of Table 3 they report that the estimated coefficient for the second term is -0.05, which is not significant, while the estimated coefficient for the first term is 2.23, which is significantly different from zero. Our analysis here suggests that the same equation can be obtained from the hybrid model without assuming a forecast smoothing.

Sticky Information and Lucas-type Noisy Information

There is another well-known type of NI structure, different from the assumption of Woodford (2003). We again consider a hybrid of two classes of information rigidities, SI and NI. However, regarding the NI part, we assume that agent i can observe the past values, $\pi_{t-1}, \pi_{t-2}, \ldots$, in period t if she can update the information in period t. This type of imperfect information assumption has long been employed in many studies since the seminal work of Lucas (1972), and recent examples include Angelotos and La'O (2009) and Crucini, Shintani, and Tsuruga (2015), among others. In principle, the true state can be revealed to the agent after any j period between $j = 1$ (the Lucas-type) and $j \to \infty$ (the Woodford-type). Therefore, in terms of the timing for agents to find out the true state, Lucas-type NI and Woodford-type NI can be viewed as two extreme cases of NI.

When SI is combined with Lucas-type NI, the average FE is given by

$$\begin{aligned}\pi_{t+h} - F_t\pi_{t+h} = &\frac{1}{1-\lambda}\left(\lambda + \frac{1-G}{G}\right)\left(F_t\pi_{t+h} - F_{t-1}\pi_{t+h}\right) \\ &- \frac{1-G}{G}\frac{1}{1-\lambda}\sum_{j=0}^{\infty}\left(-\frac{1-G}{G}\right)^j\left(\lambda + \frac{1-G}{G}\right)\left(F_{t-j-1}\pi_{t+h} - F_{t-j-2}\pi_{t+h}\right) + \nu_{t+h,t}.\end{aligned} \tag{2}$$

This equation for the average FE embeds the case of pure SI ($\lambda > 0$ and $1 - G = 0$) as well as pure NI ($\lambda = 0$ and $1 - G > 0$). As in the hybrid model with Woodford-type NI, in the case of pure SI, the FE equation reduces to equation (5) of Coibion and Gorodnichenko (2015). In the case of pure NI, however, the second term on the right-hand side of the equation does not disappear. Unless $1 - G = 0$, the average FE in period t depends on the FR in period $t - 1$ and earlier (the second term on the right-hand side of the equation), as well as that in period t (the first term). For this reason, to estimate the pure NI model, we need to extend Coibion and Gorodnichenko's regression with additional regressors of lagged FRs. Most importantly, in Lucas-type NI, a test for the null hypothesis of pure SI ($1 - G = 0$) has nontrivial power against an alternative of pure NI ($\lambda = 0$). This feature of Lucas-type NI differs from the case of Woodford-type NI discussed above because the identification of pure SI and pure NI has now become possible by extending Coibion and Gorodnichenko's regression.[5]

2.2 Single-equation Estimation for Average Forecast Errors

We now revisit the analysis of Coibion and Gorodnichenko (2015) using the same dataset: the forecasts of U.S. inflation from the SPF (Online Appendix B contains the descriptive statistics of the data). The observation period extends from 1970:1Q to 2014:2Q. The inflation rate is based on the GDP/GNP deflator.

In Coibion and Gorodnichenko's regression based on FE equation (1), we cannot conduct a test for the null hypothesis of SI since such a test has no power against an alternative of Woodford-type NI. However, as discussed in Section 2.1, we can extend Coibion and Gorodnichenko's regression using FE equation (2) and conduct a test against an alternative hypothesis of Lucas-type NI. For this reason, here we consider only the latter type of NI and determine if the data are more favorable to SI or NI as a sole candidate of the information rigidity.

Specifically, we can test the null hypothesis of pure SI (H_0: $1 - G = 0$) against an alternative of pure NI ($H_1 : \lambda = 0$) by examining whether the coefficients on the FR in period $t - 1$ and earlier are all zero in equation (2). Here, we do not impose a parameter restriction among coefficients on the FR. In Sections 4 and 5, however, we directly estimate the two structural parameters, $1 - G$ and λ, by imposing theoretical restrictions of (2) in the regression.

Our strategy is to employ the two-stage least squares (2SLS) method to estimate FE equation (2) with a constant term and to examine the significance of the estimated coefficients on lagged FRs. In the SPF, we can observe FR in the previous periods,

[5] The comparison of equations (1) and (2) shows that the latter serves as a nested model in a reduced-form regression. The model of Woodford-type NI corresponds to the case in which the coefficients on the FRs in period $t - 2$ and earlier are all zero.

$F_{t-j}\pi_{t+h} - F_{t-j-1}\pi_{t+h}$, only from $j = 0$ to $3 - h$ for a certain h. This data limitation leads to bias for the estimates because the omitted variables are likely to be correlated with the explanatory variables. To circumvent the omitted-variable bias problem, we utilize instruments: $\hat{\nu}_{t-j}$ and $\hat{\nu}_{oil,t-j}$ for $j = 0$ to $3 - h$, where $\hat{\nu}_t$ and $\hat{\nu}_{oil,t}$ represent estimated shocks to inflation and change in oil prices, respectively, obtained from the regressions of $\pi_t = \rho\pi_{t-1} + \nu_t$ and $\pi_{oil,t} = \rho_{oil}\pi_{oil,t-1} + \nu_{oil,t}$. Because these instrumental variables are unexpected shocks in periods from $t-3+h$ to t, they are uncorrelated with the omitted variables that consist of the forecast revisions in period $t-4+h$ or earlier, whereas they are correlated with the explanatory variables.[6]

The new estimate of FE equation (2) is given by

$$
\begin{aligned}
\pi_t - F_t\pi_t = \quad & \underset{(0.176)}{-\,0.143} && \underset{(0.632)}{+\,1.564}\,(F_t\pi_t - F_{t-1}\pi_t) \\
& \underset{(1.220)}{-1.952}\,(F_{t-1}\pi_t - F_{t-2}\pi_t) && \underset{(1.185)}{-\,0.651}\,(F_{t-2}\pi_t - F_{t-3}\pi_t) \\
& \underset{(1.561)}{+4.057}\,(F_{t-3}\pi_t - F_{t-4}\pi_t) && +\,residuals,
\end{aligned}
$$

where the numbers in parentheses are heteroskedasticity and autocorrelation consistent standard errors. Here, we report the result of $h = 0$ because we can use the largest number of regressors. The coefficient on the FR in period t is positive and significant at the five percent level. This coefficient implies that $\lambda = 0.61$ for pure SI and $G = 0.39$ for pure NI. The Wald test statistic for the zero restrictions on all the coefficients on lagged FRs is 6.93. Since the statistic is lower than 7.81, the critical value at the five percent

[6]Coibion and Gorodnichenko (2015) use the forecasts of the inflation rate over the next four quarters (from $h = 0$ to 3). Here, we focus on the single horizon forecast using a particular h, because the number of available regressors decreases with h increases, which complicates the estimation of FE equation (2). Moreover, it is convenient in deriving equations on cross-sectional variance, which we discuss in the next section. Note also that the same data for oil prices have been also used to obtain results in Table 3 in Coibion and Gorodnichenko (2015).

significance level, the null hypothesis of pure SI ($1 - G = 0$) is not rejected against an alternative of pure Lucas-type NI ($\lambda = 0$).[7]

3 Cross-sectional Variance of Forecasts

3.1 Models for Cross-sectional Variance of Forecasts

Coibion and Gorodnichenko (2012) have pointed out that the dependence of disagreements among agents on aggregate shocks is a useful feature in distinguishing NI from SI. We now revisit this claim and derive an expression for the cross-sectional variance (CV) based on our hybrid model, which is useful for identifying the source of information rigidity. We follow Coibion and Gorodnichenko (2012) and employ CV given by

$$\begin{aligned} V_t\pi_{t+h} &= (1-\lambda)\sum_{j=0}^{\infty}\lambda^j \mathrm{Var}_{t,t-j}\left(F_{i,t-j}\pi_{t+h} - F_t\pi_{t+h}\right) \\ &= (1-\lambda)\sum_{j=0}^{\infty}\lambda^j \mathrm{Var}_{t,t-j}\left[(F_{i,t-j}\pi_{t+h} - F_{t-j}\pi_{t+h}) + (F_{t-j}\pi_{t+h} - F_t\pi_{t+h})\right], \end{aligned}$$

as a proxy for the disagreement of forecasts, where $\mathrm{Var}_{t,t-j}(\cdot)$ denotes the cross-sectional variance of forecasts in period t for agents who last update their forecasts in period $t-j$.

Online Appendix A shows that, depending on the type of NI, our hybrid model yields the following CV equation:

$$V_t\pi_{t+h} = \begin{cases} \lambda V_{t-1}\pi_{t+h} + \frac{\lambda}{1-\lambda}(F_t\pi_{t+h} - F_{t-1}\pi_{t+h})^2 + \frac{(1-\lambda)\rho^{2h}G^2}{1-(1-G)^2\rho^2}\sigma_\omega^2 & \text{(Woodford-type NI)} \\ \lambda V_{t-1}\pi_{t+h} + \frac{\lambda}{1-\lambda}(F_t\pi_{t+h} - F_{t-1}\pi_{t+h})^2 + (1-\lambda)\rho^{2h}G^2\sigma_\omega^2 & \text{(Lucas-type NI).} \end{cases} \tag{3}$$

This equation suggests that CV depends both on the variance of forecasts made in the previous period and the squared forecast revisions, $(F_t\pi_{t+h} - F_{t-1}\pi_{t+h})^2$, with common coefficients for both types of NI. This result is consistent with Coibion and Gorodnichenko

[7]However, the hypothesis is rejected at the ten percent significance level. See Online Appendix B for the additional analyses for the robustness of this result.

(2012), Andrade and Le Bihan (2013), and Hur and Kim (2016), who explain that CV is state dependent in the SI model with $\lambda > 0$. The coefficients on two regressors, lagged CV and squared FR, can be used to test the null hypothesis of pure NI ($\lambda = 0$) against an alternative of pure SI ($1 - G = 0$).

With the same regression, switching the null and alternative hypotheses is also possible. As shown in Online Appendix A, Kalman gain G is a decreasing function of the variance of noise σ_ω and as σ_ω approaches zero, G approaches one. This fact implies that the third term in (3) disappears when the noise is small, regardless of the type of NI. For this reason, testing for the zero restriction on the intercept term is equivalent to testing the null hypothesis of pure SI ($1 - G = 0$) against an alternative of pure NI ($\lambda = 0$).

3.2 Single-equation Estimation for Cross-sectional Variance

We use the same dataset as before. To obtain the CV of inflation forecasts, in each quarter, we collect each professional's inflation forecasts and calculate their cross-sectional sample variance after dropping the top and bottom one percent of samples. We estimate CV equation (3) using the OLS without imposing a parameter restriction between the two coefficients on the lagged CV and squared FR. Since the number of regressor does not change for any forecast horizon in the CV equation, in what follows, we simply set $h = 1$. Our estimate is given by

$$\begin{array}{lllll} V_t\pi_{t+1} = & 0.463 & +0.489 V_{t-1}\pi_{t+1} & +0.678(F_t\pi_{t+1} - F_{t-1}\pi_{t+1})^2 & + residuals, \\ & (0.098) & (0.060) & (0.151) & \end{array}$$

where the numbers in parentheses are heteroskedasticity and autocorrelation consistent standard errors. We obtain the Wald test statistic of 119.6 for the restriction that the coefficients on the lagged CV and squared FR are zero. Since the statistic is greater than 9.21, the critical value at the one percent level, the null hypothesis of pure NI ($\lambda = 0$)

is rejected against an alternative of pure SI ($1 - G = 0$). Online Appendix B provides the details. We also find that the constant term is significantly different from zero at the one percent level. Thus, the null hypothesis of pure SI ($1 - G = 0$) is also rejected against an alternative of pure NI ($\lambda = 0$).

4 Joint Estimation of Pure Models of Information Rigidities

In the previous section, we noted that the CV equation is useful to identify the source of information rigidity. Given the fact that the CV contains additional information, it seems natural to combine the CV equation with FE equation (1) or (2) in the estimation of the model. The joint estimation of the two equations is helpful, not only from the perspective of identification but also from the efficiency in estimating the structural parameters. In this section, we jointly estimate the FE and CV equations for pure models of information rigidities (i.e., either pure SI or pure NI). We then employ the nonnested GMM test proposed by Smith (1992) to test the null hypothesis of one of the pure models of information rigidities against the other.

4.1 Estimation

To evaluate the empirical performance of pure models of information rigidities, we jointly estimate FE equation (1) (or (2)) and CV equation (3) by applying the generalized method of moments (GMM) to the orthogonality condition for two equations. In particular, for the pure SI model, the restriction $1 - G = 0$ is imposed on two equations. In contrast, for the pure Woodford-type NI model, we use equation (1) for the average FE and impose $\lambda = 0$ on two equations. Likewise, for the pure Lucas-type NI model, we use equation (2) for the average FE and impose $\lambda = 0$ on two equations.

Recall that we used 2SLS in the estimation of the FE equation for the pure Lucas-type NI model to circumvent the omitted-variable bias in Section 2. For the purpose of applying the nonnested GMM test, we use the common set of instruments to estimate the FE equation for all the three classes of pure models of information rigidities.[8] For the CV equation, the regressors, the lagged CV and squared FR, correspond to instruments in the GMM. We also impose parameter restrictions on λ and $1 - G$ so that their values satisfy the theoretical requirements of $0 \leq \lambda < 1$ and $0 < G \leq 1$.[9]

Columns (1) to (3) in Table 1 show the estimation results.[10] In the pure SI model, the estimate of λ is 0.51 and significantly different from zero. The point estimate implies that agents update their information set every six months on average. Our estimate of λ is close to 0.54, the value reported by Coibion and Gorodnichenko (2015). In the pure NI models, the estimate of $1-G$ is similar and significantly different from both zero and one. The value is 0.54 and 0.45 in Woodford-type NI and Lucas-type NI, respectively, while Coibion and Gorodnichenko (2015) report the value of 0.54 for $1 - G$ in Woodford-type NI.

[8]Note that this set of instruments is also valid in the estimation for the other classes of pure models of information rigidities.

[9]Specifically, we impose parameter restrictions using the following reparameterization. First, we introduce parameter λ^* to replace λ with $1 - \exp(-\lambda^{*2})$. With this transformation, the range of λ becomes $0 \leq \lambda < 1$ for $-\infty < \lambda^* < \infty$. No information stickiness ($\lambda = 0$) corresponds to the case of $\lambda^* = 0$. Second, we introduce parameter G^* to replace G with $\exp(-G^{*2})$. With this transformation, the range of G becomes $0 < G \leq 1$ for $-\infty < G^* < \infty$. Information noise is absent ($1 - G = 0$) when $G^* = 0$. See Online Appendix B for the estimation results without the parameter restrictions.

[10]While the parameters we estimate are λ^* and G^*, we report the values of λ and $1 - G$ in the table because the latter values contain clearer economic meanings. Figures in square brackets represent the 95% confidence intervals. We calculate the point estimates and 95% confidence intervals as follows. Denote the point estimates for λ^* and G^* by $\hat{\lambda}^*$ and $\hat{G}^*$, respectively. Then, the point estimates for λ and $1-G$ are $1-\exp(-\hat{\lambda}^{*2})$ and $1-\exp(-\hat{G}^{*2})$. Further, denote the heteroskedastic and autocorrelation consistent standard errors for λ^* and G^* by σ_λ and σ_G, respectively. Then, the 95% confidence intervals λ and $1-G$ are calculated as $1-\exp(-(\hat{\lambda}^* \pm 2\sigma_\lambda)^2)$ and $1-\exp(-(\hat{G}^* \pm 2\sigma_G)^2)$, respectively. The lower end of confidence intervals becomes zero for λ or $1 - G$ when λ^* or G^* takes zero within its confidence interval.

4.2 Relative Performance of the Pure SI, Pure Woodford-type NI, and Pure Lucas-type NI Models

In the joint estimation of the FE and CV equations, neither pure SI nor pure NI nests the other model. Thus, we cannot test for the null hypothesis of a pure model of information rigidity simply by imposing some restriction on the coefficients of some other model, as we did in Sections 2 and 3. For this reason, we follow Smith's (1992) approach and employ a nonnested GMM test to compare two competing models, say, models A and B. Under the null hypothesis of model B against the alternative hypothesis of model A, the Cox-type statistic asymptotically follows normal distribution $N(0,1)$.[11] The null hypothesis of model A can be also considered by switching the order of two models.

Table 1 also shows the results of the nonnested GMM test. The figures in the table indicate the p-value, based on the Cox-type statistic for the null hypothesis of model B against an alternative hypothesis of model A. Because there are three pure models of information rigidities, we compute $6(= 3 \times 2)$ statistics. The table shows that the null hypothesis of pure SI is not rejected (column (1)), while both the null hypotheses of two types of pure NI are always rejected (columns (2)(3)). Thus, the pure SI model seems to be preferable to the pure NI models, according to the results of the nonnested test. At the same time, however, the J test for the overidentifying restrictions is rejected in all three models (p-value is shown in the table). This result motivates us to investigate the hybrid model that allows for the combination of SI and NI.

[11]Each of the models, A and B, is estimated by the GMM using moment functions $E[g_A(w_t, \alpha)] = 0$ and $E[g_B(w_t, \beta)] = 0$, where $g_A(w_t, \alpha)$ and $g_B(w_t, \beta)$ are $k_A \times 1$ and $k_B \times 1$ function vectors, respectively; α and β are $p_A \times 1$ and $p_B \times 1$ parameter vectors, respectively; and w_t is a vector of observable variables including instruments z_t. Here, we use the same instrumental variables z_t and the same observations for models A and B. Then, the Cox-type statistic is given by $C_T(B|A)/\hat{\omega}_B$ where $C_T(B|A) = \hat{g}'_{A,T}\hat{W}_B\sqrt{T}\hat{g}_{B,T}$, $\hat{g}_{A,T}$ equals $g_{A,T}(\hat{\alpha})$, $\hat{\alpha}$ is the GMM estimator of α in model A ($\hat{g}'_{B,T}$ is defined similarly), T is the number of observations, and $\hat{W}_B$ represents a consistent positive semi-definite estimator of the weight matrix in model B, and $\hat{\omega}_B^2 = \hat{g}'_{A,T}\hat{W}_B\hat{g}_{A,T} - \hat{g}'_{A,T}\hat{W}_B\hat{G}_B\left(\hat{G}'_B\hat{W}_B\hat{G}_B\right)^{-1}\hat{G}'_B\hat{W}_B\hat{g}_{A,T}$, where $\hat{G}_B$ represents the Jacobian matrix in model B evaluated by $\hat{\beta}$. Further, we set the A matrix in Smith (1992) at $W_{A0}^{-1}W_B$, where W_{A0} is the limit of $\hat{W}_A$ when model B is correct.

5 Joint Estimation of Hybrid Models of Information Rigidities

In this section, we allow for the combination of NI and SI and simultaneously estimate FE and CV equations of the hybrid model. We then evaluate the relative performance of hybrid models based on the model selection criterion proposed by Andrews (1999).

5.1 Estimation

As for the hybrid model of SI and Woodford-type NI, we jointly estimate the FE and CV equations given by (1) and (3) by employing the GMM. For each equation, the instruments correspond to regressors. As for the hybrid model of the SI and Lucas-type NI, we jointly estimate the FE and CV equations given by (2) and (3). As before, we use instruments $\hat{\nu}_{t-j}$ and $\hat{\nu}_{oil,t-j}$ for $j = 0, 1, 2$, in FE equation (2). As for the CV equation (3), the instruments correspond to regressors.

Columns (1) and (2) in Table 2 show the estimation results for the hybrid model where SI is combined with Woodford-type and Lucas-type NI, respectively. In the hybrid model of SI and Woodford-type NI, the estimated value of λ is 0.42 and is significantly different from zero. This result provides support for SI. In contrast, evidence of Woodford-type NI is rather weak since the point estimate of $1 - G$ is zero, namely, it is on the lower boundary of the parameter range. When we turn to the hybrid model of SI and Lucas-type NI, the estimate of λ is 0.43, which is similar to the one obtained with the hybrid model of SI and Woodford-type NI. The estimate of $1 - G$ is 0.11. Although this value is not significantly different from zero, the estimate is not on the boundary of parameter range. Moreover, the J test shows that the overidentifying restriction is not rejected.

Let us now examine the robustness of our results using different forecast horizons h.[12] While our benchmark estimation focuses on the case of $h = 1$, we can conduct a similar

[12]In Online Appendix B, we provide further estimation results by using different equations and/or instrument variables.

joint estimation by pooling equations for different h's from zero to two. Specifically, we use three equations for the FE equation and three equations for the CV equation, which differ in h ($h = 0, 1, 2$). When the hybrid model is based on the combination of SI and Woodford-type NI, the FE equation is given by equation (1). Instruments correspond to regressors of the equation, irrespective of h. When the hybrid model is based on SI and Lucas-type NI, the FE equation is given by equation (2). The number of regressors differs, depending on h owing to data availability. The explanatory variables we can use in the SPF data are $F_{t-j-1}\pi_{t+h} - F_{t-j-2}\pi_{t+h}$ from $j = -1$ to $2 - h$, and the instrument variables are $\hat{\nu}_{t-j}$ and $\hat{\nu}_{oil,t-j}$ for $j = 0$ to $3 - h$. The CV equation is given by equation (3), where instruments correspond to regressors, for both models.

Columns (3) and (4) of Table 2 show that the parameter estimates do not change much. For the hybrid model of SI and Lucas-type NI, the estimates of λ and $1 - G$ are about 0.4 and 0.2, respectively, suggesting a large degree of information stickiness and a small degree of information noise. It should be noted that the estimate of $1 - G$ is now significantly different from zero, showing the presence of Lucas-type NI. However, the overidentifying restriction is rejected by the J test for the hybrid model of SI and Lucas-type NI. For the hybrid model of SI and Woodford-type NI, the estimate of $1 - G$ is zero, suggesting the absence of Woodford-type NI.

5.2 Relative Performance of Pure and Hybrid Models of Information Rigidities

Which model performs the best in explaining the FE and CV equations together? To answer this question, in the bottom two rows of Tables 1 and 2, we report the GMM-BIC, the model selection criterion proposed by Andrews (1999).[13] The GMM-BIC is defined

[13] We can also apply the nonnested GMM test again and we report the results in Online Appendix B. However, repeating the same test too many times should be avoided because the results may be subject to a multiple testing problem. As the number of tests increases, it becomes more likely that the null hypothesis will be rejected at some point, even if the null hypothesis is correct. If we want to compare all the pairs in five models (the pure SI model, the pure Woodford-type NI model, the pure Lucas-type NI model, the hybrid model of SI and Woodford-type NI, and the hybrid model of SI and Lucas-type

by $J - (|c| - p)\log(T)$, where J is the J test statistic for overidentifying restrictions, and $|c|$, p, and T represent the number of moment conditions, the number of parameters, and the number of observations, respectively. The model with the lowest GMM-BIC should be selected as the best model.

Comparing the GMM-BIC in Tables 1 and 2, we find that the hybrid models of SI and NI (both the Woodford- and Lucas-types) yield smaller values than the pure models of either SI or NI. Therefore, the hybrid model is preferable to pure models of information rigidities. Between the two types of hybrid models of SI and NI, the one with Lucas-type NI yields the lower value (i.e., -20.6) than the one with Woodford-type NI. This result seems reasonable when we consider the fact that respondents of the SPF are professionals rather than households. It is more likely that they have easier access to the data and would learn the value of π_{t-1} in period t. Columns (3) and (4) of Table 2 show the robustness of our results. When we pool equations for different h's from zero to two, we find that GMM-BIC of the hybrid model of SI and Lucas-type NI is lower than that of the hybrid model of SI and Woodford-type NI.

6 Concluding Remarks

For the purpose of identifying the source of information rigidities in the analysis of Coibion and Gorodnichenko (2015), we have constructed a hybrid model of SI and NI and derived a very simple form showing the cross-sectional average forecast errors and the cross-sectional variance of inflation forecasts. We find that the hybrid model of SI and Lucas-type NI is useful in explaining the actual expectation formation process.

References

[1] Andrade, Philippe, and Hervé Le Bihan. 2013. "Inattentive Professional Forecasters." *Journal of Monetary Economics*, 60(8), 967–982.

NI), we need to conduct tests for $20 (= 5 \times 4)$ times.

[2] Andrews, Donald, W. K. 1999. "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation." *Econometrica*, 67(3), 543–564.

[3] Angeletos, George-Marios, Zhen Huo, and Karthik A. Sastry. 2020. "Imperfect Macroeconomic Expectations: Evidence and Theory." In *NBER Macroeconomic Annual 2020*, forthcoming.

[4] Angeletos, George-Marios, and Jennifer La'O. 2009. "Incomplete Information, Higher-order Beliefs and Price Inertia." *Journal of Monetary Economics*, 56(S), S19–S37.

[5] Armantier, Olivier, Wilbert van der Klaauw, Scott Nelson, Giorgio Topa, and Basit Zafar. 2016. "The Price Is Right: Belief Updating in a Price Information Experiment." *Review of Economics and Statistics*, 98(3), 503–523.

[6] Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer. 2020. "Overreaction in Macroeconomic Expectations." *American Economic Review*, 110(9), 2748–2782.

[7] Branch, William A. 2007. "Sticky Information and Model Uncertainty in Survey Data on Inflation Expectations." *Journal of Economic Dynamics and Control*, 31(1), 245–276.

[8] Broer, Tobias, and Alexandre N. Kohlhas. 2020. "Forecaster (Mis-)Behavior." Working Paper.

[9] Coibion, Olivier, and Yuriy Gorodnichenko. 2012. "What Can Survey Forecasts Tell Us about Information Rigidities?" *Journal of Political Economy*, 120(1), 116–159.

[10] Coibion, Olivier, and Yuriy Gorodnichenko. 2015. "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts." *American Economic Review*, 105(8), 2644–2678.

[11] Crucini, Mario J., Mototsugu Shintani, and Takayuki Tsuruga. 2010. "Accounting for Persistence and Volatility of Good-Level Real Exchange Rates: The Role of Sticky Information." *Journal of International Economics*, 81(1), 48–60.

[12] Crucini, Mario J., Mototsugu Shintani, and Takayuki Tsuruga. 2015. "Noisy Information, Distance and Law of One Price Dynamics across US Cities." *Journal of Monetary Economics*, 74(C), 52–66.

[13] Fuhrer, Jeff. 2018. "Intrinsic Expectations Persistence: Evidence from Professional and Household Survey Expectations." Working Papers 18-9, Federal Reserve Bank of Boston.

[14] Hur, Joonyoung, and Insu Kim. 2016. "Information Rigidities in Survey Data: Evidence from Dispersions in Forecasts and Forecast Revisions." *Economics Letters*, 142(C), 10–14.

[15] Kohlhas, Alexandre and Ansgar Walther. 2021. "Asymmetric Attention." *American Economic Review*, forthcoming.

[16] Lucas Jr., Robert. 1972. "Expectations and the Neutrality of Money." *Journal of Economic Theory*, 4(2), 103–124.

[17] Mankiw, N. Gregory, and Ricardo Reis. 2002. "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve." *Quarterly Journal of Economics*, 117(4), 1295–1328.

[18] Mankiw, N. Gregory, Ricardo Reis, and Justin Wolfers. 2004. "Disagreement about Inflation Expectations." In *NBER Macroeconomics Annual 2003*, Vol. 18, edited by Mark Gertler and Kenneth Rogoff, 209–248. Cambridge, MA: MIT Press.

[19] Smith, Richard J. 1992. "Non-Nested Tests for Competing Models Estimated by Generalized Method of Moments." *Econometrica*, 60(4), 973–980.

[20] Woodford, Michael. 2003. “Imperfect Common Knowledge and the Effects of Monetary Policy.” In *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, edited by Philippe Aghion, Roman Frydman, Joseph Stiglitz, and Michael Woodford, 25–58. Princeton: Princeton University Press.

Table 1: Joint Estimation for the Average Forecast Errors and Variance: Pure Models of Information Rigidities

	(1) Pure model of SI	(2) Pure model of Woodford-NI	(3) Pure model of Lucas-NI
λ	0.510	–	–
	[0.453, 0.566]		
$1-G$	–	0.536	0.450
		[0.374, 0.68]	[0.262, 0.628]
c_1	-0.059	-0.046	-0.104
	[-0.243, 0.125]	[-0.248, 0.156]	[-0.292, 0.084]
c_2	–	0.856	0.856
		[0.68, 1.032]	[0.68, 1.032]
# of obs	172	172	172
# of moments	10	10	10
# of params	2	3	3
J test	0.0000	0.0002	0.0002
Nonnested test against H_1:			
Pure SI	–	0.002	0.005
Pure Woodford-NI	0.458	–	0.000
Pure Lucas-NI	0.446	0.000	–
GMM-BIC	-1.89	-7.52	-7.16

Notes: SI represents the sticky information, while Woodford-NI and Lucas-NI represent the Woodford-type noisy information and Lucas-type noisy information, respectively. The coefficients, c_1 and c_2, represent intercepts for the equations of average forecast errors (FE_t) and cross-sectional variance (CV_t), respectively. The instrumental variables used for the FE_t equation are $\hat{\nu}_{t-j}$, $\hat{\nu}_{oil,t-j}$ for $j = 0, 1, 2$, while those used for the CV_t equation are CV_{t-1} and squared FR_t, where FR_t represents the forecast revision. Figures in square brackets show the 95 percent confidence intervals. The J test shows the p-value for the test of overidentifying restrictions. The nonnested test shows the p-value, based on Smith (1992), for the null hypothesis of a pure model of information rigidity against another type of model. GMM-BIC indicates the model selection criterion based on Andrews (1999), where smaller values are preferable.

Table 2: Joint Estimation for the Average Forecast Errors and Variance: Hybrid Models of Information Rigidities

	(1) Hybrid model of SI/Woodford-NI	(2) Hybrid model of SI/Lucas-NI	(3) Hybrid model of SI/Woodford-NI	(4) Hybrid model of SI/Lucas-NI
h	1	1	0,1,2	0,1,2
λ	0.423	0.433	0.411	0.434
	[0.355, 0.491]	[0.363, 0.502]	[0.362, 0.461]	[0.381, 0.486]
$1-G$	0	0.114	0	0.174
	[0, 0]	[0, 0.689]	[0, 0]	[0.02, 0.414]
c_1 $(h=0)$	–	–	-0.004	-0.125
			[-0.164, 0.156]	[-0.293, 0.043]
c_1 $(h=1)$	-0.032	-0.074	-0.042	-0.088
	[-0.224, 0.16]	[-0.268, 0.12]	[-0.234, 0.15]	[-0.274, 0.098]
c_1 $(h=2)$	–	–	-0.094	-0.134
			[-0.326, 0.138]	[-0.352, 0.084]
c_2 $(h=0)$	–	–	0.500	0.461
			[0.33, 0.67]	[0.287, 0.635]
c_2 $(h=1)$	0.460	0.454	0.454	0.436
	[0.31, 0.61]	[0.302, 0.606]	[0.308, 0.6]	[0.288, 0.584]
c_2 $(h=2)$	–	–	0.661	0.632
			[0.435, 0.887]	[0.406, 0.858]
# of obs	172	172	167	167
# of moments	6	10	18	30
# of params	4	4	8	8
J test	0.583	0.114	0.614	0.000
GMM-BIC	-8.346	-20.611	-42.094	-60.668

Notes: See Table 1 for the notations. Here, h indicates a forecast horizon (the unit is a quarter). The instrumental variables used for the FE_t equation from the hybrid model of SI and Woodford-type NI are FR_t and FR_{t-1}, irrespective of h. Those from the hybrid model of SI and Lucas-type NI are $\hat{\nu}_{t-j}$ and $\hat{\nu}_{oil,t-j}$ for $j = 0$ to $3-h$. The instrumental variables used for the CV_t equation are CV_{t-1} and squared FR_t for all the specifications.

Appendix for "Identifying the Source of Information Rigidities in the Expectations Formation Process"

Mototsugu Shintani* Kozo Ueda†

2021

A Model Details

A.1 Average Forecast Errors

A.1.1 Sticky Information and Woodford-type Noisy Information

We introduce the hybrid model which combines the sticky information and noisy information. We let an aggregate variable x_t, such as the inflation rate π_t , follow a stationary AR(1) process that is given by $x_t = \rho x_{t-1} + \nu_t$, where $|\rho| < 1$ and ν_t represents an i.i.d. normal shock with mean zero. As for information noise, we follow Woodford (2003) and assume that agents cannot observe the past actual value x_{t-1} (in the main test, it is the inflation rate π_{t-1}) in period t. An agent i can instead receive an individual signal x_{it} in period t, where $x_{it} = x_t + \omega_{it}$ and ω_{it} is the mean-zero normal noise, which is i.i.d. across time t and agent i with variance σ_ω^2. Regarding sticky information, we assume that agent i can update her information set and revise her forecasts with probability $1 - \lambda$, where $0 \leq \lambda < 1$. The agent i who has an opportunity to revise her forecasts $\mathcal{F}_{it} x_t$ in period t can gather all past information $(x_{it}, x_{it-1}, x_{it-2}, \cdots)$ up to t. She thus revises her forecasts $\mathcal{F}_{it} x_t$ based on $\mathcal{F}_{it-1} x_t$ as if she revised her forecasts every period up to $t - 1$ to form $\mathcal{F}_{it-1} x_t$.

The average of forecasts is given by

$$F_t x_{t+h} = (1 - \lambda) \sum_{j=0}^{\infty} \lambda^j \mathcal{F}_{t-j} x_{t+h}, \tag{A.1}$$

where $\mathcal{F}_t$ denotes average forecasts among the agents who revise their forecasts as $\mathcal{F}_{it}$ in period

*University of Tokyo (Email: shintani@e.u-tokyo.ac.jp)
†Waseda University (Email: kozo.ueda@waseda.jp).

t:

$$\begin{aligned}\mathcal{F}_{it}x_{t+h} &= \rho^h \mathcal{F}_{it}x_t = \rho^h G y_{it} + \rho^h(1-G)\mathcal{F}_{it-1}x_t \\ &= \rho^h G y_{it} + (1-G)\mathcal{F}_{it-1}x_{t+h},\end{aligned}$$

which yields

$$\begin{aligned}\mathcal{F}_t x_{t+h} &= \rho^h G x_t + (1-G)\mathcal{F}_{t-1}x_{t+h} \\ &= G x_{t+h} + (1-G)\mathcal{F}_{t-1}x_{t+h} - G\nu_{t+h,t}.\end{aligned}$$

Here, Kalman gain G depends on the degree of information noise σ_ω and the size of the aggregate shock σ_ν as

$$\begin{aligned}G &= P/(P+\sigma_\omega^2) \\ P &= \rho^2\sigma_\omega^2 P/(P+\sigma_\omega^2) + \sigma_\nu^2.\end{aligned}$$

This outcome implies that G equals one when $\sigma_\omega = 0$ (information is not noisy) and that G decreases as σ_ω increases.

We have

$$x_{t+h} - \mathcal{F}_t x_{t+h} = (1-G)(x_{t+h} - \mathcal{F}_{t-1}x_{t+h}) + G\nu_{t+h,t}. \tag{A.2}$$

Note that

$$\begin{aligned}F_{t-1}x_{t+h} &= (1-\lambda)\sum_{j=0}^{\infty}\lambda^j \mathcal{F}_{t-1-j}x_{t+h} \\ &= (1-\lambda)\sum_{j=1}^{\infty}\lambda^{j-1}\mathcal{F}_{t-j}x_{t+h}.\end{aligned}$$

$$F_t x_{t+h} - \lambda F_{t-1}x_{t+h} = (1-\lambda)\mathcal{F}_t x_{t+h}. \tag{A.3}$$

Using equations (A.2) and (A.3), we have

$$x_{t+h} - \frac{F_t x_{t+h} - \lambda F_{t-1}x_{t+h}}{1-\lambda} = (1-G)\left(x_{t+h} - \frac{F_{t-1}x_{t+h} - \lambda F_{t-2}x_{t+h}}{1-\lambda}\right) + G\nu_{t+h,t}$$

$$x_{t+h} - F_t x_{t+h} = \frac{1-(1-\lambda)G}{(1-\lambda)G}(F_t x_{t+h} - F_{t-1}x_{t+h}) - \frac{(1-G)\lambda}{(1-\lambda)G}(F_{t-1}x_{t+h} - F_{t-2}x_{t+h}) + \nu_{t+h,t}. \tag{A.4}$$

They embed both the case of only sticky information, $G = 1$, and the case of only noisy information, $\lambda = 0$. Unless information is both sticky and noisy, the second term on the right-hand side would not appear.

A.1.2 Sticky Information and Lucas-type Noisy Information

As for information noise, we now follow Lucas (1972) and assume that agents can observe the past actual value x_{t-1} in period t. In terms of sticky information, we assume that agent i who has an opportunity to revise her forecasts $\mathcal{F}_{it}x_t$ in period t can gather all past information, including x_{t-1}.

The average of forecasts is given by

$$F_t x_{t+h} = (1-\lambda)\sum_{j=0}^{\infty}\lambda^j \mathcal{F}_{t-j}x_{t+h}, \tag{A.5}$$

where $\mathcal{F}_t$ denotes average forecasts among the agents who revise their forecasts as $\mathcal{F}_{it}$ in period t:

$$\begin{aligned}
\mathcal{F}_{it}x_t &= G(y_{it} - \rho x_{t-1}) + \rho x_{t-1}, \\
\mathcal{F}_{t}x_t &= G(x_t - \rho x_{t-1}) + \rho x_{t-1}, \\
\mathcal{F}_{t}x_{t+h} &= G(\rho^h x_t - \rho^{h+1} x_{t-1}) + \rho^{h+1} x_{t-1} \\
&= G(x_{t+h} - \nu_{t+h,t}) + \rho^{h+1}(1-G)x_{t-1},
\end{aligned}$$

where G represents the Kalman gain (again, G equals one when information is not noisy (i.e., $\sigma_\omega = 0$)) and $\nu_{t+h,t} = \sum_{j=1}^{h}\rho^{h-j}\nu_{t+j}$. We further have

$$\mathcal{F}_{t-1}x_{t+h} = G(x_{t+h} - \nu_{t+h,t-1}) + \rho^{h+2}(1-G)x_{t-2},$$

and

$$\begin{aligned}
\mathcal{F}_{t}x_{t+h} - \mathcal{F}_{t-1}x_{t+h} &= G(\nu_{t+h,t-1} - \nu_{t+h,t}) + \rho^{h+1}(1-G)\left(x_{t-1} - \rho x_{t-2}\right) \\
&= G\rho^h \nu_t + \rho^{h+1}(1-G)\nu_{t-1},
\end{aligned}$$

where $\nu_{t+h,t-1} - \nu_{t+h,t} = \rho^h \nu_t$. This result leads to

$$\begin{aligned}
\nu_t &= \frac{1}{G\rho^h}\left(\mathcal{F}_{t}x_{t+h} - \mathcal{F}_{t-1}x_{t+h}\right) - \rho\frac{1-G}{G}\nu_{t-1} \\
&= \sum_{j=0}^{\infty}\left(-\rho\frac{1-G}{G}\right)^j \frac{1}{G\rho^{h+j}}\left(\mathcal{F}_{t-j}x_{t+h} - \mathcal{F}_{t-j-1}x_{t+h}\right),
\end{aligned} \tag{A.6}$$

if $(1-G)/G < 1$.

Because

$$\begin{aligned} F_{t-1}x_{t+h} &= (1-\lambda)\sum_{j=0}^{\infty}\lambda^j \mathcal{F}_{t-1-j}x_{t+h} \\ &= (1-\lambda)\sum_{j=1}^{\infty}\lambda^{j-1}\mathcal{F}_{t-j}x_{t+h}, \end{aligned}$$

we have

$$F_t x_{t+h} - \lambda F_{t-1}x_{t+h} = (1-\lambda)\mathcal{F}_t x_{t+h}. \tag{A.7}$$

Using this, we rewrite equation (A.6) as

$$\nu_t = \frac{1}{G(1-\lambda)\rho^h}\left(F_t x_{t+h} - (1+\lambda)F_{t-1}x_{t+h} + \lambda F_{t-2}x_{t+h}\right) - \rho\frac{1-G}{G}\nu_{t-1} \tag{A.8}$$

$$= \sum_{j=0}^{\infty}\left(-\rho\frac{1-G}{G}\right)^j \frac{1}{G(1-\lambda)\rho^{h+j}}\left(F_{t-j}x_{t+h} - (1+\lambda)F_{t-j-1}x_{t+h} + \lambda F_{t-j-2}x_{t+h}\right). \tag{A.9}$$

Using

$$\begin{aligned} \mathcal{F}_t x_{t+h} &= G(x_{t+h} - \nu_{t+h,t}) + \rho^{h+1}(1-G)x_{t-1} \\ &= G(x_{t+h} - \nu_{t+h,t}) + (1-G)(x_{t+h} - \nu_{t+h,t-1}) \\ &= x_{t+h} + G\rho^h\nu_t - \nu_{t+h,t-1}, \end{aligned}$$

we have

$$\begin{aligned} F_t x_{t+h} - \lambda F_{t-1}x_{t+h} &= (1-\lambda)\left\{x_{t+h} + G\rho^h\nu_t - \nu_{t+h,t-1}\right\} \\ &= (1-\lambda)\left\{x_{t+h} + G\rho^h\nu_t - \nu_{t+h,t} - \rho^h\nu_t\right\}, \end{aligned}$$

or

$$x_{t+h} - F_t x_{t+h} = \frac{\lambda}{1-\lambda}(F_t x_{t+h} - F_{t-1}x_{t+h}) + (1-G)\rho^h\nu_t + \nu_{t+h,t},$$

that is,

$$\begin{aligned} x_{t+h} - F_t x_{t+h} = {} & \frac{\lambda}{1-\lambda}(F_t x_{t+h} - F_{t-1}x_{t+h}) \\ & + \frac{1-G}{G(1-\lambda)}\left(F_t x_{t+h} - (1+\lambda)F_{t-1}x_{t+h} + \lambda F_{t-2}x_{t+h}\right) - \rho^{h+1}\frac{(1-G)^2}{G}\nu_{t-1} + \nu_{t+h,t}. \end{aligned} \tag{A.10}$$

It becomes

$$x_{t+h} - F_t x_{t+h} = \frac{\lambda}{1-\lambda}(F_t x_{t+h} - F_{t-1}x_{t+h}) \\ + \frac{1-G}{G}\sum_{j=0}^{\infty}\left(-\frac{1-G}{G}\right)^j \frac{1}{1-\lambda}\left(F_{t-j}x_{t+h} - (1+\lambda)F_{t-j-1}x_{t+h} + \lambda F_{t-j-2}x_{t+h}\right) + \nu_{t+h,t}, \tag{A.11}$$

and then

$$\begin{aligned} x_{t+h} - F_t x_{t+h} = & \frac{\lambda}{1-\lambda}(F_t x_{t+h} - F_{t-1}x_{t+h}) \\ & + \frac{1-G}{G}\sum_{j=0}^{\infty}\left(-\frac{1-G}{G}\right)^j \frac{1}{1-\lambda}\left(F_{t-j}x_{t+h} - F_{t-j-1}x_{t+h}\right) \\ & - \frac{1-G}{G}\sum_{j=0}^{\infty}\left(-\frac{1-G}{G}\right)^j \frac{\lambda}{1-\lambda}\left(F_{t-j-1}x_{t+h} - F_{t-j-2}x_{t+h}\right) \\ & + \nu_{t+h,t}, \end{aligned}$$

or

$$x_{t+h} - F_t x_{t+h} = \frac{1}{1-\lambda}\left(\lambda + \frac{1-G}{G}\right)(F_t x_{t+h} - F_{t-1}x_{t+h}) \\ - \frac{1-G}{G}\frac{1}{1-\lambda}\sum_{j=0}^{\infty}\left(-\frac{1-G}{G}\right)^j \left(\lambda + \frac{1-G}{G}\right)(F_{t-j-1}x_{t+h} - F_{t-j-2}x_{t+h}) + \nu_{t+h,t}. \tag{A.12}$$

The coefficients on forecast revisions depend on both λ and G.

A.2 Cross-sectional Variance of Forecasts

A.2.1 Sticky Information and Woodford-type Noisy Information

The variance of forecasts is given by

$$\begin{aligned} V_t x_{t+h} &= (1-\lambda)\sum_{j=0}^{\infty}\lambda^j \mathrm{Var}_{t,t-j}\left(F_{i,t-j}x_{t+h} - F_t x_{t+h}\right) \\ &= (1-\lambda)\sum_{j=0}^{\infty}\lambda^j \mathrm{Var}_{t,t-j}\left((F_{i,t-j}x_{t+h} - F_{t-j}x_{t+h}) + (F_{t-j}x_{t+h} - F_t x_{t+h})\right). \end{aligned} \tag{A.13}$$

Regarding the first term, we define $V_t^N x_{t+h}$ as

$$
\begin{aligned}
V_t^N x_{t+h} &\equiv (1-\lambda) \sum_{j=0}^{\infty} \lambda^j \mathrm{Var}_{t,t-j} \left(\bar{F}_{i,t-j} x_{t+h} - \bar{F}_{t-j} x_{t+h} \right) \\
&= (1-\lambda) \sum_{j=0}^{\infty} \lambda^j \mathrm{Var}_{t,t-j} \Big(\rho^{h+j} G \omega_{i,t-j} + (1-G) \left(\bar{F}_{i,t-j-1} x_{t+h} - \bar{F}_{t-j-1} x_{t+h} \right) \Big) \\
&= (1-\lambda) \sum_{j=0}^{\infty} \lambda^j \left(\rho^{2h+2j} G^2 \sigma_\omega^2 \right) + (1-G)^2 V_{t-1}^N x_{t+h} \\
&= \frac{(1-\lambda)\rho^{2h}}{1-\lambda\rho^2} G^2 \sigma_\omega^2 + (1-G)^2 V_{t-1}^N x_{t+h}.
\end{aligned}
$$

Regarding the second term, we define $V_t^S x_{t+h}$ as

$$
\begin{aligned}
V_t^S x_{t+h} \equiv& (1-\lambda) \sum_{j=0}^{\infty} \lambda^j \mathrm{Var}_{t,t-j} \left(\bar{F}_{t-j} x_{t+h} - F_t x_{t+h} \right) \\
=& (1-\lambda) \sum_{j=0}^{\infty} \lambda^j \left(\bar{F}_{t-j} x_{t+h} - F_t x_{t+h} \right)^2 \\
=& (1-\lambda) \sum_{j=1}^{\infty} \lambda^j \left(\bar{F}_{t-j-1} x_{t+h} - F_{t-1} x_{t+h} - \left(F_t x_{t+h} - F_{t-1} x_{t+h} \right) \right)^2 \\
& + (1-\lambda) \left(\bar{F}_t x_{t+h} - F_t x_{t+h} \right)^2 \\
=& \lambda (1-\lambda) \sum_{j=0}^{\infty} \lambda^j \left(\bar{F}_{t-j-1} x_{t+h} - F_{t-1} x_{t+h} \right)^2 + \lambda (F_t x_{t+h} - F_{t-1} x_{t+h})^2 \\
& + (1-\lambda) \left(\frac{F_t x_{t+h} - \lambda F_{t-1} x_{t+h}}{1-\lambda} - F_t x_{t+h} \right)^2 \\
=& \lambda V_{t-1}^S x_{t+h} + \frac{\lambda}{1-\lambda} (F_t x_{t+h} - F_{t-1} x_{t+h})^2
\end{aligned}
$$

using equation (A.3). Note that the interaction term of the first and second terms in equation (A.13) is zero because the first term is determined essentially only by the history of idiosyncratic shocks $\omega_{i,t-j}$, while the second term is their aggregation.

Inserting these two equations to equation (A.13), we have

$$
V_t x_{t+h} = V_t^N x_{t+h} + V_t^S x_{t+h}, \tag{A.14}
$$

where

$$V_t^N x_{t+h} = \frac{(1-\lambda)\rho^{2h}}{1-\lambda\rho^2} G^2\sigma_\omega^2 + (1-G)^2 V_{t-1}^N x_{t+h}, \tag{A.15}$$

$$V_t^S x_{t+h} = \lambda V_{t-1}^S x_{t+h} + \frac{\lambda}{1-\lambda}(F_t x_{t+h} - F_{t-1}x_{t+h})^2. \tag{A.16}$$

The variance can be also written as follows:

$$V_t x_{t+h} = \frac{(1-\lambda)\rho^{2h}}{1-\lambda\rho^2} G^2\sigma_\omega^2 + \frac{\lambda}{1-\lambda}(F_t x_{t+h} - F_{t-1}x_{t+h})^2 + \varepsilon_{t-1}, \tag{A.17}$$

where $\varepsilon_{t-1} = (1-G)^2 V_{t-1}^N x_{t+h} + \lambda V_{t-1}^S x_{t+h}$. They embed both the case of only sticky information, $G = 1$, and the case of only noisy information, $\lambda = 0$.

The above equation can be further simplified in the following way. Equation (A.15) can be written as

$$\begin{aligned}
V_t^N x_{t+h} &= \frac{(1-\lambda)\rho^{2h}}{1-\lambda\rho^2} G^2\sigma_\omega^2 + (1-G)^2 \left\{ \frac{(1-\lambda)\rho^{2(h+1)}}{1-\lambda\rho^2} G^2\sigma_\omega^2 + (1-G)^2 V_{t-2}^N x_{t+h} \right\} \\
&= \cdots = \frac{(1-\lambda)\rho^{2h}}{1-\lambda\rho^2} G^2\sigma_\omega^2 \left\{ 1 + (1-G)^2\rho^2 + \cdots \right\} \\
&= \frac{(1-\lambda)\rho^{2h}}{1-\lambda\rho^2} \frac{G^2}{1-(1-G)^2\rho^2}\sigma_\omega^2,
\end{aligned}$$

and, thus, $V_t^N \pi_{t+h}$ converges to a constant. Hence, we have

$$\begin{aligned}
V_t x_{t+h} &= V^N x_{t+h} + V_t^S x_{t+h} \\
&= \frac{(1-\lambda)\rho^{2h}}{1-\lambda\rho^2} \frac{G^2}{1-(1-G)^2\rho^2}\sigma_\omega^2 + \lambda V_{t-1}^S x_{t+h} + \frac{\lambda}{1-\lambda}(F_t x_{t+h} - F_{t-1}x_{t+h})^2 \\
&= \lambda \left\{ \frac{(1-\lambda)\rho^{2(h+1)}}{1-\lambda\rho^2} \frac{G^2}{1-(1-G)^2\rho^2}\sigma_\omega^2 + V_{t-1}^S x_{t+h} \right\} + \frac{\lambda}{1-\lambda}(F_t x_{t+h} - F_{t-1}x_{t+h})^2 \\
&\quad + (1-\lambda\rho^2)\frac{(1-\lambda)\rho^{2h}}{1-\lambda\rho^2} \frac{G^2}{1-(1-G)^2\rho^2}\sigma_\omega^2 \\
&= \lambda V_{t-1} x_{t+h} + \frac{\lambda}{1-\lambda}(F_t x_{t+h} - F_{t-1}x_{t+h})^2 + \frac{(1-\lambda)\rho^{2h}G^2}{1-(1-G)^2\rho^2}\sigma_\omega^2.
\end{aligned} \tag{A.18}$$

This result suggests that the disagreements of forecasts depend on the past disagreements, the constant term (σ_ω^2), and the revision of forecasts.

A.2.2 Sticky Information and Lucas-type Noisy Information

We have

$$
\begin{aligned}
\bar{F}_{i,t-j}x_{t+h} - F_t x_{t+h} &= \left(\bar{F}_{i,t-j}x_{t+h} - \bar{F}_{t-j}x_{t+h}\right) + \left(\bar{F}_{t-j}x_{t+h} - F_t x_{t+h}\right) \\
&= \rho^{h+j} G \omega_{i,t-j} + \left(\bar{F}_{t-j}x_{t+h} - F_t x_{t+h}\right).
\end{aligned}
$$

Thus, the variance of forecasts is written as

$$
\begin{aligned}
V_t x_{t+h} &= (1-\lambda)\sum_{j=0}^{\infty} \lambda^j \mathrm{Var}_{t,t-j}\left(\bar{F}_{i,t-j}x_{t+h} - F_t x_{t+h}\right) \\
&= (1-\lambda)\sum_{j=0}^{\infty} \lambda^j \left\{\left(\rho^{h+j}G\right)^2 \sigma_\omega^2 + \left(\bar{F}_{t-j}x_{t+h} - F_t x_{t+h}\right)^2\right\} \\
&= (1-\lambda)\sum_{j=0}^{\infty} \lambda^j \left(\rho^{h+j}G\right)^2 \sigma_\omega^2 \\
&+ (1-\lambda)\sum_{j=0}^{\infty} \lambda^j \left\{\left(\bar{F}_{t-j}x_{t+h}\right)^2 - 2\left(\bar{F}_{t-j}x_{t+h}\right)\left(F_t x_{t+h}\right) + \left(F_t x_{t+h}\right)^2\right\} \\
&= (1-\lambda)\sum_{j=0}^{\infty} \lambda^j \left(\rho^{h+j}G\right)^2 \sigma_\omega^2 + (1-\lambda)\sum_{j=0}^{\infty} \lambda^j \left(\bar{F}_{t-j}x_{t+h}\right)^2 - \left(F_t x_{t+h}\right)^2.
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
V_{t-1} x_{t+h} &= (1-\lambda)\sum_{j=0}^{\infty} \lambda^j \left\{\left(\rho^{h+1+j}G\right)^2 \sigma_\omega^2 + \left(\bar{F}_{t-1-j}x_{t+h} - F_{t-1} x_{t+h}\right)^2\right\} \\
&= (1-\lambda)\sum_{j=1}^{\infty} \lambda^{j-1} \left(\rho^{h+j}G\right)^2 \sigma_\omega^2 + (1-\lambda)\sum_{j=1}^{\infty} \lambda^j \left(\bar{F}_{t-j}x_{t+h}\right)^2 - \left(F_{t-1} x_{t+h}\right)^2.
\end{aligned}
$$

Thus, we have

$$
V_t x_{t+h} - \lambda V_{t-1} x_{t+h} = (1-\lambda)\left(\rho^h G\right)^2 \sigma_\omega^2 + (1-\lambda)\left(\bar{F}_t x_{t+h}\right)^2 - \left(F_t x_{t+h}\right)^2 + \lambda\left(F_{t-1} x_{t+h}\right)^2.
$$

Using equation (A.7), we have

$$
\begin{aligned}
V_t x_{t+h} - \lambda V_{t-1} x_{t+h} &= (1-\lambda)\left(\rho^h G\right)^2 \sigma_\omega^2 + (1-\lambda)\left(\frac{F_t x_{t+h} - \lambda F_{t-1} x_{t+h}}{1-\lambda}\right)^2 - \left(F_t x_{t+h}\right)^2 + \lambda\left(F_{t-1} x_{t+h}\right)^2 \\
&= (1-\lambda)\left(\rho^h G\right)^2 \sigma_\omega^2 + \frac{\lambda}{1-\lambda}\left(F_t x_{t+h} - F_{t-1} x_{t+h}\right)^2,
\end{aligned}
$$

which yields

$$V_t x_{t+h} = \lambda V_{t-1} x_{t+h} + \frac{\lambda}{1-\lambda} \left(F_t x_{t+h} - F_{t-1} x_{t+h}\right)^2 + (1-\lambda) \left(\rho^h G\right)^2 \sigma_\omega^2. \tag{A.19}$$

Importantly, this disagreement equation based on the hybrid model of SI and Lucas-type NI is expressed in the same form as that based on the hybrid of SI and Woodford-type NI provided in equation (A.18).

B Estimation

B.1 Data

Table B.1 shows the basic statistics of the SPF data we use.

B.2 Estimation Results

B.2.1 Single-equation Estimation for Average Forecast Errors

Table B.2 shows the estimation results of FE equation (1) or (2) that are shown in the main text. Depending on the type of NI models and the forecast horizon of h, we use different equations and explanatory variables. In columns (1), (2), (3), (6), and (7), we assume Lucas-type NI (i.e., the pure Lucas-type NI model or the hybrid model of SI and Lucas-type NI). Specifically, in column (1), we set $h = 1$ where explanatory variables are $F_t\pi_{t+1} - F_{t-1}\pi_{t+1}$, $F_{t-1}\pi_{t+1} - F_{t-2}\pi_{t+1}$, and $F_{t-2}\pi_{t+1} - F_{t-3}\pi_{t+1}$, and instrumental variables are $\hat{\nu}_{t-j}$ and $\hat{\nu}_{oil,t-j}$ for $j = 0, 1, 2$ in employing the 2SLS (i.e., the number of instrumental variables is six, which is denoted by IV(6) in the table).

The last row represents the Wald test statistics for the zero restrictions on coefficients on the forecast revisions in periods $t-1$ and $t-2$. This statistic can be used to test the null hypothesis of pure SI (H_0: $G = 1$) against an alternative of pure Lucas-type NI ($H_1 : \lambda = 0$).

Alternatively, we use instrumental variables of $\hat{\nu}_{t-j}$ for $j = 0, 1, 2$ (IV(3) in column (2)) or conduct the OLS (column (3)). If we estimate equation (2) in the main text using the finite number of regressors, the omitted variables lead to a bias. This is the reason we use the instrumental variables, but this bias may be small if one of information rigidities is small ($G \simeq 1$ or $\lambda \simeq 0$) because the coefficient on the past FR approaches zero quickly as j increases. If this is the case, we do not need to use instrumental variables. We run a regression of FE on current and past FRs, namely, $F_t\pi_{t+1} - F_{t-1}\pi_{t+1}$, $F_{t-1}\pi_{t+1} - F_{t-2}\pi_{t+1}$, and $F_{t-2}\pi_{t+1} - F_{t-3}\pi_{t+1}$; that is, we employ the OLS.

In column (4), we assume either SI or Woodford-type NI. In this case, the FE equation does not embed FRs in period $t-1$ and $t-2$, and thus, we cannot test the null hypothesis of pure SI (H_0: $G=1$) against an alternative of pure Woodford-type NI ($H_1 : \lambda = 0$). Thus, the Wald test statistic is not reported in the table. In column (5), we assume the hybrid model of SI and Woodford-type NI. Columns (6) and (7) show the estimation results when we use a different forecast horizon of $h=0$ and 2, respectively.

B.2.2 Single-equation Estimation for Cross-sectional Variance

Table B.3 shows the estimation results of the CV equation (3) in the main text (equation (A.18) in the appendix). In column (1), we set $h=1$ and employ the OLS.

The CV equation can be expressed and estimated differently. In columns (2) and (3), we estimate equation (A.17), rather than equation (A.18), by employing the 2SLS with the instrumental variables of $\hat{\nu}_t$, $\hat{\nu}_t^2$, $\hat{\nu}_{oil,t}$, and $\hat{\nu}_{oil,t}^2$ (i.e., the number of instrumental variables is four, which is denoted by IV(4) in the table). In column (3), we estimate the same equation as column (2) but with the instrumental variables of $\hat{\nu}_t$ and $\hat{\nu}_t^2$ (denoted by IV(2)). Note that in equation (A.17), the third and fourth terms on the right-hand side of the equation are not only unobservable but also correlated with the second term because both depend on past shocks for x_t, that is, ν_{t-i} for $i=1,2,\cdots$. Thus, if we regress the equation with the OLS using $(F_t x_{t+h} - F_{t-1}x_{t+h})^2$ as an independent variable, the estimate is biased. However, $(F_t x_{t+h} - F_{t-1}x_{t+h})^2$ depends on today's shock ν_t, while the third and fourth terms do not.[1] Therefore, if we use today's new information as instrumental variables, we expect to obtain an unbiased coefficient for the term of $(F_t x_{t+h} - F_{t-1}x_{t+h})^2$. Considering that the FR is squared in equation (A.17), we use $\hat{\nu}_t^2$ as well as $\hat{\nu}_t$.

Columns (4) and (5) examine the robustness of our results to changes in the measurement of the CV. We exclude the top and bottom 2.5 percent of samples in column (4) and the samples in which the inflation forecasts exceed 30 percent in their absolute term in column (5). Finally,

[1] To be precise, it is shown in the following way. Note that

$$\begin{aligned} \digamma_t x_{t+h} &= \rho^h G x_t + (1-G)\digamma_{t-1}x_{t+h} \\ &= \rho^h G \nu_t + \rho^{h+1} G x_{t-1} + (1-G)\digamma_{t-1}x_{t+h} \end{aligned}$$

and

$$\begin{aligned} F_t x_{t+h} &= (1-\lambda)\sum_{j=0}^{\infty}\lambda^j \digamma_{t-j}x_{t+h} = (1-\lambda)\digamma_t x_{t+h} + (1-\lambda)\sum_{j=1}^{\infty}\lambda^j \digamma_{t-j}x_{t+h} \\ &= (1-\lambda)\rho^h G\nu_t + (1-\lambda)\left\{\rho^{h+1}Gx_{t-1} + (1-G)\digamma_{t-1}x_{t+h} + \sum_{j=1}^{\infty}\lambda^j \digamma_{t-j}x_{t+h}\right\}. \end{aligned}$$

The second term is determined at $t-1$ or earlier, and is thus independent of ν_t. Thus, it is clear that $(F_t x_{t+h} - F_{t-1}x_{t+h})^2$ depends on ν_t unless $\lambda = 1$, $G=0$, or $\rho = 0$.

in columns (6) and (7), we estimate the CV equation for $h = 0$ and 2, respectively.

B.2.3 Joint Estimation of Pure Models of Information Rigidities

We consider one of pure models of information rigidities, SI or one type of NI. The results are shown in Table B.4. In columns (1) to (3), we use different instrumental variables for the FE equation, that is, $F_t\pi_{t+1} - F_{t-1}\pi_{t+1}$, $F_{t-1}\pi_{t+1} - F_{t-2}\pi_{t+1}$, and $F_{t-2}\pi_{t+1} - F_{t-3}\pi_{t+1}$. They are the same variables as those appearing on the right-hand side of the FE equation based on the pure Lucas-type NI model, which implies that we estimate the FE equation by the OLS. As we stated in Section B.2.1, the bias in the estimates will be small if one of information rigidities is small ($G \to 1$ or $\lambda \to 0$) when using the OLS.

In columns (4) to (6), we estimate the model without imposing parameter restrictions on λ and G, which allows them to take below zero and above one.

The estimation results do not change much. In the pure SI model, the estimate of λ is around 0.5 and significantly different from zero. In pure NI models, the estimate of $1 - G$ is around 0.3 to 0.5, depending on the models and significantly different from zero. According to the J test, the validity of overidentifying restrictions is rejected in all the models. Furthermore, the GMM-BIC based on pure models of information rigidities is higher, and thus, worse, than that based on the hybrid model of SI and Lucas-type NI.

B.2.4 Joint Estimation of Hybrid Models of Information Rigidities

We check the robustness of the estimation results when we use a different forecast horizon: $h = 0$ or 2. Table B.5 shows the robustness of our main results. The estimate of λ is around 0.4 and significantly different from zero. The estimate of $1 - G$ is 0 (corner solution) when based on Woodford-type NI, but positive when based on Lucas-type NI. In particular, it is 0.3, which is significantly positive when $h = 0$. However, in this model, the validity of overidentifying restrictions is rejected.

As in Section B.2.3, we also check the robustness of the estimation results when we adopt a different estimation strategy. Table B.6 shows the results, where columns (1) to (3) indicate the case of using different equations and/or instrumental variables and columns (4) and (5) show the case of not imposing restrictions on λ and G.

Specifically, in column (1), the hybrid model of SI and Woodford-type NI is estimated based on the CV equation (A.17), rather than (A.18). As explained in Section B.2.2, we estimate this equation by using $\hat{\nu}_t$, $\hat{\nu}_t^2$, $\hat{\nu}_{oil,t}$, and $\hat{\nu}_{oil,t}^2$ as instrumental variables. Column (1) shows that the value of λ is 0.54, while the value of $1 - G$ is zero. Furthermore, the intercept of the CV equation, c_2, is significantly different from zero, which is inconsistent with $G = 1$.

Columns (2) and (3) show the estimation results for the hybrid model of SI and Lucas-type NI. In column (2), we reduce the number of instrumental variables for the FE equation from six to three by using $\hat{\nu}_{t-j}$ $(j = 0, 1, 2)$. The estimation results do not change much, compared to those shown in column (2) in Table 2 in the main text, which is an outcome suggesting the robustness of our results. In column (3), we estimate the FE equation using the variables on the right-hand side of the equation (i.e., OLS). In this case, the estimate of $1 - G$ reaches zero (corner solution). The intercept of the CV equation, c_2, is significantly different from zero, which is inconsistent with $G = 1$. Nevertheless, the estimate of λ is robust at around 0.4.

In Columns (4) and (5), we do not impose parameter restrictions for λ and G. Specifically, λ and G can take any values ($-\infty$ to ∞). The estimation results show that $1 - G$ becomes negative for the hybrid model of SI and Woodford-type NI. The results are almost unchanged for the hybrid model of SI and Lucas-type NI.

Last and not least, the comparison of the GMM-BIC shows that column (2) in Table 2 in the main text yields the lowest value, and thus, the best fit of all the specifications.

B.2.5 Nonnested Test

In Table B.7, we report the p-values, based on the Cox-type statistic proposed by Smith (1992), to compare five nonnested models. They are the pure SI model, the pure Woodford-type NI model, the pure Lucas-type NI model, the hybrid model of SI and Woodford-type NI, and the hybrid model of SI and Lucas-type NI. One-by-one test of five models leads to $20(= 5 \times 4)$ statistics.

For comparison, we use the same instruments. instrumental variables are $\hat{\nu}_{t-j}$ and $\hat{\nu}_{oil,t-j}$ $(j = 0, 1, 2)$ for the FE equation and CV_{t-1} and FR_t^2 for the CV equation. The estimation results show that all the models tend to be rejected by an alternative. However, it is important to note that this one-by-one test involves the so-called multiple testing problem. As the number of inferences increases, the more likely a null hypothesis would be rejected and, in turn, erroneous inferences are made.

Table B.1: Descriptive Statistics

	Horizon h	Obs	Mean	S.D.
Forecast error (FE)	0	182	0.020	1.147
$\pi_{t+h} - F_t\pi_{t+h}$	1	181	0.044	1.487
	2	180	0.030	1.685
	3	179	0.021	1.823
Forecast revision (FR)	0	181	0.023	0.663
$F_t\pi_{t+h} - F_{t-1}\pi_{t+h}$	1	181	-0.006	0.520
	2	181	0.003	0.410
	3	176	0.014	0.435
Disagreements (V)	0	182	1.420	1.631
$V_t\pi_{t+h}$	1	182	1.297	1.780
	2	182	1.367	1.702
	3	182	1.462	2.057

Table B.2: Regression of the Cross-sectional Average Forecast Error

	(1) Pure L-NI or hybrid of SI/NI IV(6)	(2) Pure L-NI or hybrid of SI/NI IV(3)	(3) Pure L-NI or hybrid of SI/NI OLS	(4) Pure SI or pure W-NI	(5) Hybrid of SI/W-NI
FR_t	-0.586	10.967	0.523	0.655	0.58
	(1.254)	(33.587)	(0.423)	(0.455)	(0.400)
FR_{t-1}	1.667	-0.287	0.417		0.540*
	(1.278)	(9.093)	(0.357)		(0.313)
FR_{t-2}	2.182	-22.479	0.212		
	(1.360)	(72.893)	(0.389)		
FR_{t-3}					
c_1	-0.065	0.488	-0.017	0.04	0.023
	(0.174)	(1.698)	(0.137)	(0.143)	(0.131)
# of obs	172	173	173	180	179
R^2			0.070	0.053	0.077
F for weak ident	2.536	0.030			
Wald for zero restr	3.889	0.127	2.418		

	(6) Pure L-NI or hybrid of SI/NI $h=0$	(7) Pure L-NI or hybrid of SI/NI $h=2$
FR_t	1.564**	1.069
	(0.632)	(1.221)
FR_{t-1}	-1.952	2.928**
	(1.220)	(1.265)
FR_{t-2}	-0.651	
	(1.185)	
FR_{t-3}	4.057***	
	(1.561)	
c_1	-0.143	-0.089
	(0.176)	(0.175)
# of obs	172	172
R^2		
F for weak ident	1.684	4.230
Wald for zero restr	6.926*	5.36**

Notes: ***, **, and * indicate that coefficients are significant at the 1 percent, 5 percent, and 10 percent level, respectively. NI represents the model of noisy information based on either Woodford (2003, W-NI) or Lucas (1972, L-NI), whereas SI represents the model of sticky information. FR_t represents forecast revisions at t and h indicates a forecast horizon (the unit is a quarter, benchmark h is one). The figures in parentheses show the heteroskedastic and autocorrelation consistent (HAC, Newey–West) standard errors. The last two rows represent the F statistics for weak identification and the Wald statistics for the hypothesis that all the coefficients on the FR in period $t-1$ and earlier are zero, respectively.

Table B.3: Regression of the Cross-sectional Variance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Omit	Omit		
	OLS	IV(4)	IV(2)	$\pm 2.5\%$	$\lvert\pi\rvert > 30\%$	$h = 0$	$h = 2$
FR_t^2	0.678***	1.746***	1.723***	0.582***	1.302**	0.800***	0.934***
	(0.151)	(0.536)	(0.544)	(0.135)	(0.544)	(0.270)	(0.340)
CV_{t-1}	0.489***			0.493***	0.288**	0.373***	0.342***
	(0.060)			(0.081)	(0.134)	(0.120)	(0.123)
c_2	0.463***	0.809***	0.819***	0.391***	0.719***	0.531***	0.733***
	(0.098)	(0.195)	(0.194)	(0.082)	(0.164)	(0.120)	(0.197)
# of obs	181	180	181	181	181	181	181
R^2	0.361			0.371	0.269	0.447	0.195
F stat for weak ident		6.467	3.609				
Wald for zero restr	119.6***			83.74***	47.95***	74.18***	18.89***

Notes: See Table 2 for the notations. FR_t^2 and CV_{t-1} represent the squared forecast revisions and the cross-sectional variance of forecasts in the previous period, respectively. In IV(4), the instrumental variables are $\hat{\nu}_t$, $\hat{\nu}_t^2$, $\hat{\nu}_{oil,t}$, and $\hat{\nu}_{oil,t}^2$. In IV(2), the instrumental variables are the first two variables used in IV(4). The Wald statistic is for the hypothesis that the coefficients on both FR_t^2 and CV_{t-1} are zero, respectively.

Table B.4: Robustness Check: Joint Estimation of Pure Models of Information Rigidities

	(1)	(2)	(3)	(4)	(5)	(6)
		Different IV			Unrestricted λ, G	
Model	Pure SI	Pure W-NI	Pure L-NI	Pure SI	Pure W-NI	Pure L-NI
λ	0.489			0.511		
	[0.433, 0.543]			[0.455, 0.567]		
$1 - G$		0.354	0.256		0.536	0.450
		[0.093, 0.639]	[0.016, 0.604]		[0.308, 0.62]	[0.36, 0.74]
c_1	-0.014	-0.069	-0.106	-0.059	-0.046	-0.104
	[-0.204, 0.176]	[-0.281, 0.143]	[-0.304, 0.092]	[-0.243, 0.125]	[-0.248, 0.156]	[-0.292, 0.084]
c_2		0.856	0.856		0.856	0.856
		[0.68, 1.032]	[0.68, 1.032]		[0.68, 1.032]	[0.68, 1.032]
# of obs	172	172	172	172	172	172
# of moments	7	7	7	10	10	10
# of params	2	3	3	2	3	3
J test	0.0000	0.0004	0.0003	0.0000	0.0002	0.0002
Nonnested test against H_1:						
Pure SI	–	0.018	0.034	–	0.002	0.005
Pure W-NI	0.016	–	0.000	0.458	–	0.000
Pure L-NI	0.007	0.000	–	0.446	0.000	–
Model comparison criteria						
GMM-BIC	7.88	-0.28	0.49	-1.89	-7.52	-7.16

Notes: SI represents the model of sticky information, while W-NI and L-NI represent the Woodford- and Lucas-type model of noisy information, respectively. The coefficients, c_1 and c_2, represent intercepts for the equations of average forecast errors (FE_t) and cross-sectional variance (CV_t), respectively. In columns (1) to (3), the instrumental variables used for the FE_t equation are FR_t, FR_{t-1}, and FR_{t-2}, where FR_t represents the forecast revision. In columns (4) to (6), the instrumental variables used for the FE_t equation are $\hat{\nu}_{t-j}$, $\hat{\nu}_{oil,t-j}$ for $j = 0, 1, 2$. The instrumental variables used for the CV_t equation are CV_{t-1} and FR_t^2 in all the columns. In columns (4) to (6), no restriction is imposed for the range of values that λ and G can take. Figures in square brackets show the 95% confidence intervals. J test shows the p-value for the test of overidentifying restrictions. The nonnested test shows the p-value, based on Smith (1992), for the null hypothesis of a pure model of information rigidity against another type of model. GMM-BIC indicates the model selection criterion based on Andrews (1999) where smaller values are preferable.

Table B.5: Robustness Check: Joint Estimation of Hybrid Models of Information Rigidities (Different Forecast Horizons)

	(1)	(2)	(3)	(4)
	Hybrid model of		Hybrid model of	
	SI/W-NI	SI/L-NI	SI/W-NI	SI/L-NI
h	0	0	2	2
λ	0.386	0.422	0.434	0.442
	[0.297, 0.474]	[0.323, 0.519]	[0.32, 0.546]	[0.316, 0.565]
$1-G$	0	0.271	0	0.001
	[0, 0]	[0.1, 0.473]	[0, 0]	[0, 1]
c_1 $(h=0)$	-0.024	-0.13		
	[-0.184, 0.136]	[-0.308, 0.048]		
c_1 $(h=2)$			-0.088	-0.159
			[-0.318, 0.142]	[-0.385, 0.067]
c_2 $(h=0)$	0.549	0.488		
	[0.341, 0.757]	[0.262, 0.714]		
c_2 $(h=2)$			0.647	0.638
			[0.387, 0.907]	[0.368, 0.908]
# of obs	172	172	172	172
# of moments	6	12	6	8
# of params	4	4	4	4
J test	0.623	0.000	0.468	0.071
GMM-BIC	-8.532	-8.687	-7.753	-11.941

Notes: See Table 4 for the notations. h indicates a forecast horizon (the unit is a quarter). The instrumental variables used for the FE_t equation from the hybrid model of SI and Woodford-type NI are FR_t and FR_{t-1} irrespective of h. Those from the hybrid model of SI and Lucas-type NI are $\hat{\nu}_{t-j}$ and $\hat{\nu}_{oil,t-j}$ for $j=0$ to $3-h$. The instrumental variables used for the CV_t equation are CV_{t-1} and FR_t^2 for all the specifications.

Table B.6: Robustness Check: Joint Estimation of Hybrid Models of Information Rigidities (Different instrumental Variables and Less Restricted Parameters)

	(1)	(2)	(3)	(4)	(5)
		Different IV		Unrestricted λ, G	
	SI/W-NI	SI/L-NI	SI/L-NI	SI/W-NI	SI/L-NI
λ	0.541	0.432	0.423	0.437	0.433
	[0.429, 0.644]	[0.362, 0.501]	[0.355, 0.491]	[0.369, 0.505]	[0.363, 0.503]
$1-G$	0	0.075	0	-0.222	0.114
	[0, 0]	[0, 0.898]	[0, 0]	[-0.76, 0.316]	[-0.338, 0.566]
c_1	0.031	-0.088	-0.043	-0.022	-0.074
	[-0.179, 0.241]	[-0.302, 0.126]	[-0.231, 0.145]	[-0.222, 0.178]	[-0.268, 0.12]
c_2	0.804	0.454	0.46	0.486	0.454
	[0.564, 1.044]	[0.302, 0.606]	[0.31, 0.61]	[0.324, 0.648]	[0.302, 0.606]
# of obs	172	172	172	172	172
# of moments	8	7	7	6	10
# of params	4	4	4	4	4
J test	0.11	0.07	0.70	0.45	0.11
GMM-BIC	-11.56	-8.42	-13.23	-9.11	-20.61

Notes: See Table 4 for the notations. See the text for the choice of the instrumental variables.

Table B.7: Nonnested GMM Test

H_0 vs H_1	Pure SI	Pure W-NI	Pure L-NI	Hybrid of SI/W-NI	Hybrid of SI/L-NI
Pure SI	–	0.458	0.446	0.000	0.000
Pure W-NI	0.002	–	0.000	0.001	0.001
Pure L-NI	0.006	0.000	–	0.001	0.001
Hybrid of SI/W-NI	0.000	0.046	0.012	–	0.001
Hybrid of SI/L-NI	0.000	0.003	0.002	0.001	–

Notes: The nonnested GMM test shows the p-value, based on Smith (1992), for the null hypothesis of a certain model H_0 (each row) against another type of model H_1 (each column). The instrumental variables used for the FE_t equation are $\hat{\nu}_{t-j}$ and $\hat{\nu}_{oil,t-j}$ $(j = 0, 1, 2)$, while those used for the CV_t equation are CV_{t-1} and FR_t^2.