

A Likelihood Analysis of Models with Information Frictions

Leonardo Melosi*

University of Pennsylvania

February 25, 2009

Abstract

This paper develops a dynamic stochastic general equilibrium model where firms are imperfectly informed. We estimate the model through likelihood-based methods and find that it can explain the highly persistent real effects of monetary disturbances that are documented by a benchmark VAR. The model of imperfect information nests a model of rational inattention where firms optimally choose the variances of signal noise, subject to an information-processing constraint. We present an econometric procedure to evaluate the predictions of this rational inattention model. Implementing this procedure delivers insights on how to improve the fit of rational inattention models.

JEL Classifications: E3, E5, C32, D8.

Keywords: Imperfect common knowledge; rational inattention; Bayesian econometrics; real effects of nominal shocks; VAR identification.

*Corresponding author: Leonardo Melosi, 160 McNeil, 3718 Locust Walk, Philadelphia, PA 19104-6297, USA, email address: lmelosi@econ.upenn.edu, mobile phone: 267-939-4165. I thank Frank Diebold, Jesús Fernández-Villaverde, Dirk Krueger, Bartosz Maćkowiak, Frank Schorfheide, and Mirko Wiederholt for very helpful comments and discussion. I am grateful for comments from participants at the Midwest Macro Meetings 2008, the 23rd Annual Congress of the European Economic Association, Upenn Macro Lunches and Upenn Econometric Lunches.

1 Introduction

This paper develops and estimates a dynamic stochastic general equilibrium (DSGE) model where agents are imperfectly informed, as in Woodford (2002). This type of model is well-suited to explaining highly persistent real effects of money and delayed effects on inflation (Woodford, 2002), which are documented by VAR studies (Christiano *et al.*, 1999, Stock and Watson, 2001, Christiano *et al.*, 2005). Furthermore, this model has another appealing feature as it nests a simple model of rational inattention where firms optimally choose what to pay attention to, subject to an information-processing constraint à la Sims (2003). Whether these models can generate sluggish real effects of nominal shocks hinges upon the parameter values that determine how informed agents are. A shortcoming of the literature is the lack of empirical guidance in selecting these parameter values. We try to counter this shortcoming by estimating these parameters through Bayesian methods.

The paper contributes to the existing literature along three dimensions. First, we show that the estimated model of imperfect information à la Woodford (2002) can account for the strongly persistent real effects of monetary disturbances that characterize the impulse response functions of a benchmark VAR. Second, we present an econometric procedure that evaluates whether the predictions of the rational inattention model are supported by the data. Third, by implementing this procedure, we gain insights into how to improve the fit of rational inattention models.

Following Woodford (2002), we assume that firms do not perfectly observe any realizations of the model variables. There are two state variables in the model: the aggregate technology and the monetary policy stance. Firms observe idiosyncratic noisy signals regarding the state variables and solve a signal extraction problem in order to keep track of the model variables. Since the signal is noisy, firms do not immediately learn the occurrence of monetary disturbances. As a result, the price level fails to adjust enough to entirely neutralize the real effects of nominal shocks (Lucas, 1973). Moreover, because of

the idiosyncratic nature of the signals, in the aftermath of a shock firms are also uncertain about what other firms know that other firms know... that other firms know about that shock. This feature of the model is termed imperfect common knowledge. When firms find it optimal to react to changes of endogenous variables (e.g., in the presence of strategic complementarity in price setting), a problem of forecasting the forecast of others of the type envisioned by Townsend (1983b) arises. This feature of the model has been shown to amplify the persistence in economic fluctuations (Townsend, 1983a, 1983b; Hellwig, 2002; Adam, 2008; Angeletos and La'O, 2008; Rondina, 2008; and Lorenzoni, forthcomingA), and in the propagation of monetary disturbances to real variables and prices (Phelps, 1970; Lucas, 1972; Woodford, 2002; Adam, 2007; Gorodnichenko, 2008; Maćkowiak and Wiederholt, 2008; Nimark, 2008; Paciello, 2008; and Lorenzoni, forthcomingB).¹

We evaluate the fit of the model with imperfect common knowledge. For this purpose, we introduce a model that deviates from the one of imperfect common knowledge in only two respects: (1) all agents are perfectly informed, and (2) firms can optimally adjust their prices only at random periods, as in Calvo (1983). The last assumption is common to a very large number of models that have been used as workhorses for monetary policy studies over the last 25 years. We fit both models to a data set that includes U.S. per capita GDP and the U.S. GDP deflator. First, we find that the model with imperfect common knowledge fits the data better than the Calvo model. Second, the model with imperfect information can largely accommodate the persistent real effects of monetary shocks implied by a benchmark VAR. Third, when we replace the mechanism of imperfect common knowledge with that of sticky prices à la Calvo, we observe that such persistence substantially drops.

We modify the model of imperfect common knowledge so as to allow firms to optimally choose the variances of signal noise given an information-processing constraint à la Sims (2003). This model of rational inattention is nested into the model with imperfect common

¹See Mankiw and Reis (2002a, 2002b, 2006, 2007), and Reis (2006a, 2006b, 2009) for models with information frictions that do not feature imperfect common knowledge but can generate sizeable persistence.

knowledge. The former model makes predictions over the variances of the signal noise. In the latter model these variances are instead structural parameters whose values are learned from the data by estimating the model. We introduce and implement an econometric procedure that allows us to assess to what extent the predictions of this simple model of rational inattention are supported by the data. We find that these predictions are rejected by the data to some extent. Moreover, this exercise delivers interesting insights on how to improve the fit of rational inattention models. In this respect, we observe that capital accumulation would be an important feature to be added to these models.

The procedure to evaluate the predictions of the model of rational inattention can be summarized in four steps. First, we sample with replacement the posterior draws for the parameters of the model with imperfect common knowledge. Second, for each sampled draw, we measure how much information firms acquire per unit of time in the model with imperfect common knowledge. Third, for each sampled draw, we solve the model of rational inattention by using the output of the second step to determine the tightness of firms' information-processing constraint. Fourth, we evaluate whether the variances of signal noise predicted by the two models are similar.

We depart from Woodford (2002) in two respects. First, our empirical strategy is likelihood-based, while Woodford (2002) calibrates the parameters of his model. Second, Woodford's model has one rather than two shocks. Having an additional shock allows us to get around the problem of stochastic singularity when we evaluate the likelihood function. Specifically, we consider a nominal shock and an aggregate technology shock.

This paper is also related to the literature of rational inattention (Sims, 2003, 2006; Luo, 2008; Paciello, 2008; Van Nieuwerburgh and Veldkamp, 2008; Woodford, 2008; and Maćkowiak and Wiederholt, forthcoming). Maćkowiak and Wiederholt (2009) introduce a model where firms optimally decide how much attention to pay to aggregate and idiosyncratic conditions, subject to a constraint on information flows. When they calibrate their model

to match the average absolute size of price changes observed in micro data, they find that nominal shocks have sizeable and persistent real effects.

The rest of the paper is organized as follows. Section 2 presents both the model with imperfect common knowledge and the model of rational inattention, as well as the Calvo model. Some features of the first two models are explored in section 3. Section 4 deals with the empirical analysis. In section 5, we conclude.

2 The models

In this section we describe three DSGE models. The first model is a model with imperfect common knowledge (henceforth, **ICK model**). In this model, information-processing frictions are modelled by assuming that firms have to solve a signal extraction problem in order to estimate the state of the aggregate technology and that of monetary policy. A feature of this model is that firms take the stochastic process of signals as given. In the second model (henceforth, **rational inattention model**) firms solve the same signal extraction problem as in the ICK model but they are allowed to optimally choose the variances of signal noise, subject to an information-processing constraint of the type used in Sims (2003). In the third model (henceforth, **Calvo model**) all agents have perfect information but they can re-optimize their prices only at random periods, as in Calvo (1983). In the first part of this section we introduce the equations common to all the models. In the remaining part of the section, we analyze the specific features of the three models.

2.1 The common structure

The economy is populated by households, final goods producers (or producers), intermediate goods firms (or firms), a financial intermediary, and a monetary authority (or central bank). Households derive utility from consumption of final goods and disutility from supplying labor

to the intermediate goods firms. Furthermore, households face a cash-in-advance (CIA) constraint. The final goods producers are perfectly competitive with a CES production function. The intermediate goods firms operate in a monopolistic competitive environment with a production function that is linear in its unique input, which is labor. Furthermore, there are two shocks: an aggregate productivity shock that affects intermediate goods firms' technology and a monetary policy shock.

At the beginning of period t , the households inherit the entire money stock of the economy, M_t . They decide how much money D_t to deposit at the financial intermediary. These deposits yield interest at rate $R_{H,t} - 1$. The financial intermediary receives household deposits and a monetary injection from the monetary authority, which it lends to final goods producers at rate $R_{F,t} - 1$. The intermediate goods firms hire labor services from households and produce their output. The firms sell their output to the final goods producers and use the proceeds to pay wages, $W_t H_t$, where W_t is the nominal hourly wage, and H_t is hours worked, and dividends, Π_t , to households. Households' cash balance increases to $M_t - D_t + W_t H_t + \Pi_t$. The CIA constraint requires that households pay for all consumption purchases with the accumulated cash balances. The producers sell the final goods to households and then pay back their loans. Finally, households receive back their deposits inclusive of interest rate and the net cash inflow of the financial intermediary as dividend Π_t^b .

2.1.1 The representative household

The representative household solves the problem:

$$\max_{\{C_t, H_t, M_{t+1}, D_t\}} \mathbb{E}_t \sum_{s=0}^{\infty} \beta^s [\ln C_{t+s} - \alpha H_{t+s}]$$

such that

$$P_t C_t \leq M_t - D_t + W_t H_t + \Pi_t \quad (1)$$

$$0 \leq D_t \quad (2)$$

$$M_{t+1} = (M_t - D_t + W_t H_t + \Pi_t - P_t C_t) + R_{H,t} D_t + \Pi_t^b \quad (3)$$

137 where C_t is the amount of the final good consumed at time t , P_t is the price of the final good
 138 at time t , and β is the discount factor.

139 **2.1.2 The technology of the intermediate goods firms**

140 Every intermediate goods firm has the same technology:

$$Y_{i,t} = A_t N_{i,t} \quad (4)$$

141 where $Y_{i,t}$ is the output produced by the firm i at time t , and $N_{i,t}$ is the labor input demanded
 142 by firm i at time t .

143 We further assume that the aggregate productivity A_t follows a random walk with drift:

$$\ln A_t = \ln a + \ln A_{t-1} + \sigma_a \varepsilon_{a,t} \quad (5)$$

144 where $\varepsilon_{a,t} \sim \mathcal{N}(0, 1)$. Finally, it turns out to be useful to define:

$$a_t \equiv \ln A_t - \ln a \cdot t \quad (6)$$

145 **2.1.3 The final goods producers**

146 The representative final goods producer combines a continuum of intermediate goods indexed
 147 by $i \in [0, 1]$ by using the CES technology:

$$Y_t = \left(\int_0^1 (Y_{i,t})^{\frac{\nu-1}{\nu}} di \right)^{\frac{\nu}{\nu-1}} \quad (7)$$

148 where the parameter ν is assumed to be strictly larger than unity.

149 The producer takes input prices P_t^i and output price P_t as given. Furthermore, it has to
 150 borrow the cash needed to pay the intermediate goods firms at rate $R_{F,t} - 1$. Hence, its cost
 151 function is $(\int P_t^i Y_{i,t} di) R_{F,t}$. Profit maximization implies that the demand for intermediate
 152 goods will be:

$$Y_{i,t} = \left(\frac{P_t^i}{P_t} \right)^{-\nu} Y_t \quad (8)$$

153 where the competitive price of the final good P_t is given by

$$P_t = \left(\int (P_t^i)^{1-\nu} di \right)^{\frac{1}{1-\nu}} \quad (9)$$

154 2.1.4 The financial intermediary

155 The financial intermediary solves the trivial problem:

$$\max_{\{L_t, D_t\}} \mathbb{E}_t \left[\sum_{s=0}^{\infty} \beta^s \frac{\Pi_{t+s}^b}{Q_{t+s+1}} \right] \quad (10)$$

st

$$\Pi_t^b = D_t + R_{F,t} L_t - R_{H,t} D_t - L_t + X_t \quad (11)$$

$$L_t \leq X_t + D_t \quad (12)$$

156 where Q_t is the time 0 value of a unit of the consumption good in period t to the representative
 157 household and $X_t = M_{t+1} - M_t$ is the monetary injection.

2.1.5 The monetary authority

The monetary authority sets the growth rate of money so as to ensure that a log-linear combination of output and price level follows an exogenous process of the following type:

$$\Delta \ln \Lambda_t = (1 - \rho_\Lambda) \Lambda^* + \rho_\Lambda \Delta \ln \Lambda_{t-1} + \sigma_\Lambda \varepsilon_{\Lambda,t} \quad (13)$$

with $\varepsilon_{\Lambda,t} \sim \mathcal{N}(0, 1)$ and

$$\ln \Lambda_t = \lambda \ln Y_t + \ln P_t \quad (14)$$

where Δ stands for the first-difference operator, the degree of smoothness in conducting monetary policy ρ_Λ is such that $\rho_\Lambda \in [0, 1)$. Λ^* is a parameter that represents the long-run average growth rate of $\ln \Lambda_t$. Moreover, the monetary policy shock $\varepsilon_{\Lambda,t}$ is assumed to be orthogonal to the productivity shock $\varepsilon_{a,t}$. Finally, it is useful to denote:

$$m_t \equiv \ln \Lambda_t - \Lambda^* \cdot t \quad (15)$$

2.2 ICK model

In the ICK model, intermediate goods firms do not face any cost when they adjust their prices. Nonetheless, they cannot observe any realizations of the model variables. Firms observe idiosyncratic noisy signals concerning the state of technology $\ln A_t$ and that of monetary policy $\ln \Lambda_t$. Therefore, they will estimate the model variables by using the history of realizations of their signals. For tractability, it is assumed that the other agents perfectly observe the past and the current realizations of the model variables.

The intermediate goods firms solve:

$$\max_{P_t^i} \mathbb{E} [\beta^t Q_t (P_t^i Y_{i,t} - W_t N_{i,t}) | \mathcal{I}_t^i], \quad \forall t \in \{1, 2, \dots\} \quad (16)$$

st.

$$Y_{i,t} = \left(\frac{P_t^i}{P_t} \right)^{-\nu} Y_t, \quad Y_{i,t} \leq A_t N_{i,t} \quad (17)$$

$$\mathcal{I}_t^i = (\{\mathbf{z}_{i,\tau}\}_{\tau=-\infty}^t, \Theta_I) \quad (18)$$

where Q_t is the time 0 value of a unit of the consumption good in period t to the representative household, which is treated as exogenous by the firm. $\mathcal{I}_t^i$ is the information set available to firm i at time t . This set contains the history of the idiosyncratic signals $\{\mathbf{z}_{i,\tau}\}_{\tau=-\infty}^t$ and the vector of model parameters Θ_I , that is

$$\Theta_I \equiv (\nu, \rho_\Lambda, \alpha, \ln a, \Lambda^*, \lambda, \beta, \sigma_\Lambda, \sigma_a, \sigma_{e_1}, \sigma_{e_2}) \quad (19)$$

It is important to emphasize that we assume that at time 0 firms are endowed with an infinite sequence of signals. This assumption simplifies the analysis. Furthermore, the equilibrium laws of motion of all model variables are assumed to be common knowledge among firms.

Firm i 's signal model is

$$\begin{bmatrix} z_{1,i,t} \\ z_{2,i,t} \end{bmatrix} = \begin{bmatrix} m_t \\ a_t \end{bmatrix} + \begin{bmatrix} e_{1,i,t} \\ e_{2,i,t} \end{bmatrix} \quad (20)$$

where $\mathbf{z}_{i,t} \equiv [z_{1,i,t}, z_{2,i,t}]'$, $\mathbf{e}_{i,t} \equiv [e_{1,i,t}, e_{2,i,t}]'$ and

$$\mathbf{e}_{i,t} \stackrel{iid}{\sim} \mathcal{N}(\mathbf{0}, \Sigma_e), \quad \Sigma_e = \begin{bmatrix} \sigma_{e_1}^2 & 0 \\ 0 & \sigma_{e_2}^2 \end{bmatrix} \quad (21)$$

Note that a_t and m_t are the state variables of the model and the signal noises $e_{1,i,t}$ and $e_{2,i,t}$

are assumed to be *iid* across firms and time.

Assuming that the two signals are orthogonal may be considered a strong assumption. After all, firms might learn about a given state variable by processing signals concerning the other state variable. We find, however, that relaxing this assumption of orthogonality of signals does not substantially affect the main predictions of the estimated model.

Finally, one should notice that, as in Woodford (2002), firms are assumed to perfectly observe neither the amount of labor hired $N_{i,t}$ nor the quantity sold $Y_{i,t}$. They are able to get information about these variables indirectly through their estimates of the state variables.

2.3 The rational inattention model

The model of rational inattention relies on three fundamental assumptions. First, information about all model variables is freely available to decision makers. Second, information needs to be processed before being used for decision-making. Third, intermediate goods firms face limitations on the amount of information they can process per unit of time. As a result, firms will optimally decide how much information they want to acquire about each variable that matters for their price-setting decisions. For tractability, it is assumed that the other agents do not face any information-processing constraints.

In full-fledged models of rational inattention (e.g., Maćkowiak and Wiederholt, forthcoming), agents optimally choose the stochastic process of signals, subject to an information-processing constraint à la Sims (2003). Unlike these models, we parametrically restrict the set of signal processes that firms can select. Specifically, we assume that firms optimally choose among signals that follow a "true state plus white noise Gaussian error" process. Hence, what firms are allowed to choose are the variances of signals in equations (20)-(21). Nevertheless, one can show that the signal process (20)-(21) is not optimal if profit function is not quadratic or λ is not equal to unity (Maćkowiak and Wiederholt, forthcoming, sections 6 and 7). We introduce these parametric restrictions for tractability. Moreover, we

assume that firms can choose the stochastic process of signals at time 0 but they cannot reconsider their decision thereafter. In section 3.2, we will show that this last assumption is not critical for our results.

At period zero, firms allocate their attention by solving:

$$\max_{\sigma_{e_1,i}, \sigma_{e_2,i}} \mathbb{E} \left[\sum_{t=1}^{\infty} \beta^t Q_t (P_{i,t}^* Y_{i,t} - W_t N_{i,t}) \mid \mathcal{I}_t^i \right], \quad (22)$$

st

$$P_{i,t}^* = \arg \max_{P_t^i} \mathbb{E} [\beta^t Q_t (P_t^i Y_{i,t} - W_t N_{i,t}) \mid \mathcal{I}_t^i] \quad (23)$$

$$Y_{i,t} = \left(\frac{P_{i,t}^*}{P_t} \right)^{-\nu} Y_t, \quad Y_{i,t} \leq A_t N_{i,t} \quad (24)$$

$$\mathcal{I}_t^i = (\{\mathbf{z}_{i,\tau}\}_{\tau=-\infty}^t, \Theta_R) \quad (25)$$

$$\begin{bmatrix} z_{1,i,t} \\ z_{2,i,t} \end{bmatrix} = \begin{bmatrix} m_t \\ a_t \end{bmatrix} + \begin{bmatrix} e_{1,i,t} \\ e_{2,i,t} \end{bmatrix} \quad (26)$$

$$\mathbf{e}_{i,t} \stackrel{iid}{\sim} \mathcal{N}(\mathbf{0}, \Sigma_e), \quad \Sigma_e = \begin{bmatrix} \sigma_{e_1,i}^2 & 0 \\ 0 & \sigma_{e_2,i}^2 \end{bmatrix} \quad (27)$$

$$\kappa_{m,i,t} + \kappa_{a,i,t} \leq \kappa, \quad \text{any } t > 0 \quad (28)$$

where Θ_R is a vector including all the parameters of the model,

$$\Theta_R \equiv (\nu, \rho_\Lambda, \alpha, \ln a, \Lambda^*, \lambda, \beta, \sigma_\Lambda, \sigma_a, \kappa) \quad (29)$$

The variables $\kappa_{m,i,t}$ and $\kappa_{a,i,t}$ denote the information flow from signal $z_{1,i,t}$ to the state of monetary policy, m_t , and that from signal $z_{2,i,t}$ to the state of technology, a_t , respectively. Moreover, the parameter κ quantifies the overall amount of information firms can process in each period. Finally, we define the vector $\mathbf{z}_{i,t} \equiv [z_{1,i,t}, z_{2,i,t}]'$.

Notice that firms have to solve two problems: a price-setting problem and a problem of how to allocate their attention between the two state variables. In the problem of allocating the attention, firms optimally choose the variances of signal noise. Notice that when firms decide how to allocate their attention, they are aware that this choice will affect the objective function (23) and in turn the optimal price-setting policy. Moreover, conditional to these variances of signal noise, rationally inattentive firms face the same price-setting problem as that in the ICK model.

The information set (25) is of the same type as that in the ICK model. Equations (26)-(27) restrict the set of signal processes that can be chosen by firms to be "true state plus white noise Gaussian error" processes. The information-processing constraint (28) sets an upper bound $\kappa \in \mathbb{R}_+$ on the overall amount of information firms can gather at any time t . We define the information flows $\kappa_{m,i,t}$ and $\kappa_{a,i,t}$ in this constraint as follows:

$$\kappa_{m,i,t} \equiv H(m_t | z_{1,i}^{t-1}) - H(m_t | z_{1,i}^t) \quad (30)$$

$$\kappa_{a,i,t} \equiv H(a_t | z_{2,i}^{t-1}) - H(a_t | z_{2,i}^t) \quad (31)$$

where $H(m_t | z_{1,i}^\tau)$ and $H(a_t | z_{2,i}^\tau)$ are the conditional entropies of the state variable m_t and a_t , given the history of signals up to time τ , $\mathbf{z}_i^\tau$. In information theory (Shannon, 1948), entropy is an axiomatic measure of conditional uncertainty about random variables (Ash, 1990). For instance, the entropy of m_t conditional to the sequence of signals $z_{1,i}^t$ is given by $\int_{-\infty}^{\infty} \log_2 [p(m_t | z_{1,i}^t)] p(m_t | z_{1,i}^t) dm_t$, where $p(m_t | z_{1,i}^t)$ is the conditional probability density function of m_t . Since all shocks and noise in the model are Gaussian, one can show that the following results hold:

$$H(m_t | z_{1,i}^\tau) \equiv \frac{1}{2} \log_2 [2\pi e \cdot \text{VAR}(m_t | z_{1,i}^\tau)] \quad (32)$$

$$H(a_t|z_{2,i}^\tau) \equiv \frac{1}{2} \log_2 [2\pi e \cdot VAR(a_t|z_{2,i}^\tau)] \quad (33)$$

See Cover and Thomas (1991). The unit of measure of these conditional entropies and consequently that of information flows $\kappa_{m,i,t}$ and $\kappa_{a,i,t}$ is 1 bit.² Moreover, as in the ICK model, we assume that the equilibrium laws of motion of all variables are common knowledge.

2.4 A sticky price model à la Calvo (1983)

In the Calvo model all agents perfectly observe the past and current realizations of the model variables. Moreover, the prices charged by each firm are re-optimized only at random periods. The key (simplifying) assumption is that the probability that a given firm will adjust its price within a particular period is independent of the state of the model, the current price charged, and how long ago it was last re-optimized. Firms that do not re-optimize index their prices at the balance-growth-path inflation rate.

We assume that only a fraction $(1 - \theta_p)$ of firms re-optimize their prices, while the remaining θ_p fraction does not reset them. The problem of the intermediate goods firms that are allowed to adjust their prices in period t is:

$$\max_{P_t^i} \mathbb{E}_t \sum_{s=0}^{\infty} \theta_p^s \beta^{t+s} Q_{t+s} (P_t^i - MC_{t+s}) \frac{Y_{i,t+s}}{P_{t+s}} \quad (34)$$

st.

$$MC_{t+s} = \frac{W_{t+s}}{A_{t+s}}, \quad Y_{i,t+s} = \left(\frac{P_t^i}{P_{t+s}} \right)^{-\nu} Y_{t+s} \quad (35)$$

where Q_{t+s} is the marginal utility of a unit of consumption at time $t + s$ in terms of the utility of the representative household at time t , and MC_{t+s} stands for the nominal marginal costs in period $t + s$. We consider only the symmetric equilibrium at which all firms will

²If we had used the natural logarithm instead of the logarithm of base two in equation (32)-(33), these quantities would have been measured in nats.

253 choose the same optimal price $P_t^i = P_t^*$. On aggregate, we have

$$P_t^{1-\nu} = \left[(1 - \theta_p) P_t^{*(1-\nu)} + \theta_p (\pi_* P_{t-1})^{1-\nu} \right] \quad (36)$$

254 where π_* is the balance-growth-path (gross) inflation rate. We denote Θ_C as the set of
 255 parameters of the Calvo model:

$$\Theta_C \equiv (\nu, \rho_\Lambda, \alpha, \ln a, \Lambda^*, \lambda, \beta, \theta_p, \sigma_\Lambda, \sigma_a) \quad (37)$$

256 3 Log-linearization and features of the models

257 All the models presented in the previous section are log-linearized before being solved. The
 258 exogenous processes (5) and (13) induce both a deterministic and a stochastic trend to all
 259 endogenous variables, except labor. We will detrend the non-stationary variables before
 260 log-linearizing the models. It is useful to define the stationary variables as follows:

$$y_t \equiv \frac{Y_t}{A_t}, \quad p_t \equiv \frac{A_t^\lambda P_t}{\Lambda_t}, \quad p_t^i = \frac{A_t^\lambda P_t^i}{\Lambda_t} \quad (38)$$

261 In order to log-linearize the models with information frictions,³ we take the following
 262 steps. First, we derive the price-setting equation by solving the intermediate goods firms'
 263 problem in both models with information frictions. Second, we transform the variables
 264 according to the definitions (38). Third, we log-linearize the resulting price-setting equation
 265 around the perfect-information symmetric steady state. Henceforth, when we refer to the
 266 three models we mean their log-linear approximations.

³How to log-linearize and solve the Calvo model is standard and hence omitted. We use the routine *gensys* developed by Sims (2002) to numerically solve this model.

3.1 Quantifying the size of information frictions in the ICK model

The following definitions turn out to be useful for evaluating the size of the information frictions in the log-linear ICK model.

Definition: *Firms' overall level of attention $\varkappa$ is the amount of information that firms process about both state variables in the unit of time.*

Definition: *Firms' allocation of attention to a given state variable is the ratio of the amount of processed information about that state variable to the overall level of attention.*

The overall level of attention $\varkappa$ is defined as $\varkappa \equiv \kappa_m + \kappa_a$, where κ_m and κ_a are computed exactly as the information flows in equations (30)-(33). The quantities $\varkappa$, κ_m and κ_a turn out not to vary across periods and firms⁴ and are all measured in bits. Moreover, the *allocation of attention* to the state of technology Υ_a can be computed as follows: $\Upsilon_a \equiv \frac{\kappa_a}{\varkappa}$.

Characterizing the parameter $\varkappa$ and Υ_a for the log-linearized ICK model requires computing the conditional variances of m_t and a_t in equations (32)-(33) for a given set of parameters Θ_I . In order to numerically pin down these variances, one has to apply the Kalman filter to the state-space model whose transition equations are given by equations (5) and (13) and the measurement equations are defined by equations (20)-(21). We can concisely represent this result through the mapping ϕ_I :

$$(\varkappa, \kappa_m, \kappa_a)' = \phi_I(\Theta_I) \quad (39)$$

We denote the pair of information flows (κ_m, κ_a) as firms' allocation of attention in the ICK model.

⁴Since firms are assumed to receive infinitely many signals at time $t = 0$, the conditional variances $VAR(m_t | z_{1,i}^\tau)$ and $VAR(a_t | z_{1,i}^\tau)$, $\tau \in \{t, t-1\}$ any $t > 0$, do not change over time. Moreover, in the ICK model, these conditional variances are the same across firms because firms face the same variances of signal noise and all shocks are Gaussian. If these variances do not change across periods and firms, neither do information flows κ_m and κ_a . See equations (32)-(33).

3.2 Some property of the rational inattention model

In the log-linear rational inattention model, firms' profit function is log-quadratic. It can be shown that when the profit function is quadratic, the optimal signal is Gaussian (Sims, 2003). This implies that the assumption we made in section 2.3 that signals follow a Gaussian process is not critical.

In section 2.3, we also assumed that firms decide their allocation of attention at time 0. They are not allowed to reconsider the allocation of attention in any subsequent periods. If firms' profit function is quadratic, this assumption does not give rise to a problem of time inconsistency of firms' policies. The reason behind this result is as follows.⁵ When firms' profit function is quadratic, it can be shown that the allocation-of-attention problem (22)-(28) turns out to be that of choosing the variances of signal noise so as to minimize the conditional variance of the profit-maximizing price under perfect information (i.e., when $\kappa \rightarrow \infty$). This conditional variance does not change over time in periods $t > 0$ because firms receive an infinite sequence of signals at time $t = 0$ and the rational inattention model is linear and Gaussian. Therefore, the objective function of the allocation-of-attention problem does not change over time, and hence, firms do not have any incentives to reconsider their allocation of attention in periods $t > 0$.

Moreover, if their profit function is quadratic, the optimal variances of signal noise can be shown to be the same across firms. Since all shocks are Gaussian and firms receive an infinite sequence of signals at time $t = 0$, the conditional variance of the profit-maximizing price under perfect information is the same for all firms. Therefore, in a quadratic-Gaussian framework, the objective function of the allocation-of-attention problem is the same across firms. Thus, every firm will find it optimal to choose the same allocation of attention. The optimal variances of signal noise will be denoted $(\sigma_{e_1}^*)^2$ and $(\sigma_{e_2}^*)^2$.

⁵A more detailed proof of this result is provided in Maćkowiak and Wiederholt (2009).

3.3 Nestedness of the ICK and the rational inattention model

For any given $\kappa \in \mathbb{R}_+$, the rational inattention model is nested within the ICK model and sets restrictions upon the variances of signal noise, $\sigma_{e_1}^2$ and $\sigma_{e_2}^2$, in the latter model. Equivalently, for given $\kappa \in \mathbb{R}_+$, the rational inattention model can be seen as casting restrictions upon firms' allocation of attention (κ_m, κ_a) in the ICK model through the mapping ϕ_I . Therefore, we can parsimoniously represent these restrictions by means of the following mapping:

$$(\kappa_m^*, \kappa_a^*)' = \phi_R(\tilde{\Theta}_R, \kappa) \quad (40)$$

where we denote the κ_m^*, κ_a^* as the information flows predicted by the rational inattention model and the set $\tilde{\Theta}_R$ as the set of parameters in Θ_R except κ . Note that $\tilde{\Theta}_R$ is a subset of Θ_I .

The following two facts are useful for removing the degree of freedom associated with assigning a value to the parameter κ . First, as showed in section 3.1, given the parameter values of the ICK model, we can quantify the overall level of attention $\varkappa$ in this model through the mapping ϕ_I . Second, when the objective function of the allocation-of-attention problem is quadratic, the information-processing constraint (28) is always binding. Therefore, we can eliminate the degree of freedom by restricting the parameter κ to be equal to firms' overall level of attention, $\varkappa$, in the ICK model. Hence we can rewrite the mapping (40) as follows:

$$(\kappa_m^*, \kappa_a^*)' = \phi_R[\tilde{\Theta}_R, \varkappa] \quad (41)$$

where $\varkappa$ is determined by the function ϕ_I in equation (39). Finally, note that the mapping ϕ_R is now a function of only the parameters in Θ_I .

3.4 Solving linear models with information frictions

A typical challenge in finding a rational expectation equilibrium (REE) in models with imperfect common knowledge is dealing with an infinite-dimensional state vector. Hence, finding an REE in the models with information frictions would require characterizing infinitely many equilibrium laws of motion (*infinite regress*). This task is clearly unmanageable. In the two models with information frictions, this problem solely arises when there is strategic complementarity in price-setting. Moreover, in these two models this issue can be elegantly resolved as in Woodford (2002) who suggests a method that can be applied to numerically solve the ICK model.

The rational inattention model is solved in four steps. First, we guess the values of the variances of signal noise, $\sigma_{e_1}^2$ and $\sigma_{e_2}^2$. Second, given this guess, we numerically characterize the law of motion of the price level exactly as we do when solving the ICK model. Third, we obtain the optimal variances of signal noise, $(\sigma_{e_1}^*)^2$ and $(\sigma_{e_2}^*)^2$, by solving the quadratic approximation of the allocation-of-attention problem in (22)-(28). Fourth, we check whether the guess made in the first step is correct, that is, whether $\|\sigma_{e_j} - \sigma_{e_j}^*\| < \varepsilon$, for $j \in \{1, 2\}$ with $\varepsilon > 0$ and small. If this criterion is not satisfied, we do another loop by setting $\sigma_{e_j} = \sigma_{e_j}^*$, for $j \in \{1, 2\}$. Otherwise, we stop.

4 Empirical analysis

This section contains the econometric analysis of the paper. We take the ICK model and the Calvo model to the data through Bayesian techniques. We do not directly estimate the rational inattention model, since obtaining a reliable approximation of posterior distributions does not turn out to be possible. Nevertheless, we present and implement an econometric procedure that formally evaluates to what extent the predictions of the rational inattention model over firms' allocation of attention are supported by the data. This exercise is inter-

esting for two reasons. First, as shown by Woodford (2002), firms' allocation of attention crucially affects the differential responsiveness of prices to different types of disturbances in models with information processing frictions. Second, this exercise can detect sources of misspecification of rational inattention models and delivers insights on how to improve the fit of models of this variety.

4.1 The data

The data are quarterly and range from the third quarter of 1954 to the fourth quarter of 2005. We use the U.S. per capita real GDP and the U.S. GDP deflator from Haver Analytics (Haver mnemonics are in italics). Per capita real GDP is obtained by dividing the nominal GDP (GDP) by the population 16 years and older ($LN16N$) and deflating using the chained-price GDP deflator ($JGDP$). The GDP deflator is given by the appropriate series ($JGDP$).

4.2 Measurement equations

Denote the U.S. per capita real GDP, and the U.S. GDP deflator as $\{GDP_t, t = 1, 2, \dots, T\}$, and $\{DEFL_t, t = 1, 2, \dots, T\}$, respectively. The measurement equations are:

$$\ln GDP_t = \hat{y}_t + a_t + \ln a \cdot t + \ln \bar{y} \quad (42)$$

$$\ln DEFL_t = \hat{p}_t + m_t - \lambda a_t + (\Lambda^* - \lambda \ln a) \cdot t + \ln \bar{p} \quad (43)$$

where the subscript $\hat{\cdot}$ means log-deviations of a variable from its perfect-information symmetric steady-state value, $\ln \bar{y}$ is the logarithm of the steady-state value of y_t , and $\ln \bar{p}$ is the logarithm of the steady-state value of p_t .

The Kalman filter can be used to evaluate the likelihood function of the models. Yet, the filter must be initialized and a distribution for the state vector in period $t = 0$ has

to be specified. As far as the vector of stationary state variables is concerned, we use their unconditional distributions. We cannot initialize the vector of non-stationary state variables (i.e. m_t, a_t) in the same manner, since their unconditional variance is not defined. We follow the approach introduced by Chang *et al.* (2007), who propose to factorize the initial distribution as $p(\mathbf{s}_{1,t})p(\mathbf{s}_{2,t})$, where $\mathbf{s}_{1,t}$ and $\mathbf{s}_{2,t}$ are the vector of stationary and non-stationary variables, respectively. They suggest setting the first component $p(\mathbf{s}_{1,t})$ equal to the unconditional distribution of $\mathbf{s}_{1,t}$, whereas the second component $p(\mathbf{s}_{2,t})$ is absorbed into the specification of the prior.

4.3 Priors for the model parameters

We use the same prior distributions for those parameters that are common across models. We fix the value of ν equal to 10. This implies a mark-up of about 11%, which is in line with what is suggested by Woodford (2003). Table 1 elicits the prior distributions for the parameters used in both the ICK model and the Calvo model.

In the ICK model, the parameter λ entirely gauges the strategic complementarity in price setting, which is measured by $(1 - \lambda^{-1})$. As shown by Woodford (2002), this crucially affects the persistence in the mechanism of shock propagation in the ICK model. Hence, we set a broad prior for this parameter in order to educe its value from the likelihood. The prior median is set at $\lambda = 6.67$ so that the model exhibits the degree of strategic complementarity suggested by Woodford (2003).

We note that, conditional to λ , we observe $\ln \Lambda_t$. Hence, the autoregressive parameter of monetary policy, ρ_Λ , the standard deviation of the monetary policy shock, σ_Λ , and the trend Λ^* are directly estimated when λ is set equal to its prior median. We center the priors for these three parameters accordingly. Furthermore, we set broad prior intervals for these parameters.

The prior of the standard deviation of the productivity shock, σ_a , is centered at 0.007.

This value is regarded as plausible by the real business cycle literature (Prescott, 1986). Moreover, we center the prior for $\ln a$ consistently with the estimated linear trend of the U.S. per capita real output.

In absolute terms, we set the priors for standard deviations of signal noise, σ_{e_1} , and σ_{e_2} , so as to ensure that signals are quite informative about the business-cycle variations of model variables.⁶ In relative terms, these prior specifications are chosen so as to make each signal equally informative about the corresponding state variable. More specifically, we want the prior median of the allocation of attention, Υ_a , to be approximately equal to 0.5. The 90% confidence interval for Υ_a is broad, ranging from 0.16 to 0.88. The rationale of such a large confidence range is that allocation of attention is a crucial parameter affecting the differential responsiveness of prices to different types of disturbances. Thus, we aim at learning the value of Υ_a from the likelihood.

The discount factor, β , is well known in the literature, and hence we set its prior standard deviation relatively small. The prior confidence interval for β includes 0.99, which is a plausible discount factor when the model periods are interpreted as quarters (Woodford, 2002). The prior for the Calvo parameter θ_p is centered at 0.67, implying an average duration of price contracts of three quarters. This value is regarded as consistent with the survey evidence discussed in Blinder *et al.* (1998). The parameter α is not identifiable, since we do not have hours worked among our observables.

4.4 Posteriors for parameters in the ICK and the Calvo model

Given the priors and the likelihood functions implied by the models, a closed-form solution for the posterior distributions for parameters cannot be derived. However, we are able to evaluate the posteriors numerically through the random-walk Metropolis-Hastings algorithm.

⁶We achieve that by setting the prior medians of the coherences between the process of the state variables, in first difference, and their corresponding signals such that these are not smaller than 0.50 at business-cycle frequencies (3-5 years). The coherence ranges from 0 to 1 and measures the degree to which two stationary stochastic processes are jointly influenced by cycles of a given frequency (Hamilton, 1994).

How these procedures apply to macro DSGE models is exhaustively documented by An and Schorfheide (2007). We generate 1,000,000 draws from the posteriors. The posterior medians and 95% confidence intervals are shown in table 2. The posterior median of the Calvo parameter θ_p implies that firms reset their prices about every four years. This frequency of price adjustments is implausible, according to the existing microeconomic analyses on price changes. Nonetheless, this result is not surprising. In fact, it is well-known that small-scale DSGE models with sticky prices à la Calvo can match the persistence of the macro data only with price contracts of very long duration (Bils and Klenow, 2004). We might fix this problem by setting a tighter prior for the Calvo parameter, but we find that this would seriously undermine the fit of the Calvo model.

The coefficient $(1 - \lambda^{-1})$ controls the degree of strategic complementarity in price settings. As shown by Woodford (2002), this coefficient is very important, since it affects the persistence of the IRFs of output and price level to structural shocks. The prior median of strategic complementarity $(1 - \lambda^{-1})$ was set at 0.84. Hence, Bayesian updating points toward a lower strategic complementarity than what is conjectured in the prior. This tends to reduce the persistence in the mechanism of shock propagation.

Moreover, the posterior median of the signal-to-noise ratio regarding the state of monetary policy, $\sigma_{e_1}/\sigma_\Lambda$, is large relative to that associated with the state of technology, σ_{e_2}/σ_a . These estimates imply that the signal regarding the state of technology conveys more information than the signal concerning the state of monetary policy. In table 2, we also report the posterior moments for the allocation of attention (i.e. $\varkappa$, κ_m , and κ_a) in the ICK model. We find that firms can process up to 0.27 bits per quarter. About 84% of the overall level of attention is allocated to the state of technology. In every quarter, firms acquire 0.04 bits of information about the state of monetary policy and 0.23 bits about the state of technology.

4.5 Evaluating the fit of the ICK model

In this section, we assess how accurately the ICK model fits the data relative to the Calvo model. Moreover, we introduce a VAR that can be considered a benchmark because it fits the data better than these two DSGE models. We then evaluate the ICK model and the Calvo model in terms of their capability of accommodating features of the IRFs implied by the identified VAR.

4.5.1 MDD-based comparisons

From a Bayesian perspective, the issue of whether the ICK model fits the data better than the Calvo model can be addressed by comparing the marginal data densities (MDDs) of these two models (Kass and Raftery, 1995 and An and Schorfheide, 2007). Let us denote the ICK model and the Calvo model with $\mathcal{M}_I$ and $\mathcal{M}_C$, respectively. The data used for estimation are denoted by $\tilde{Y} = \{\ln GDP_t, \ln DEFL_t, t = 1, \dots, T\}$. The MDDs for the ICK model, $P(\tilde{Y}|\mathcal{M}_I)$, and the Calvo model, $P(\tilde{Y}|\mathcal{M}_C)$, are defined as:

$$P(\tilde{Y}|\mathcal{M}_I) = \int \mathcal{L}(\Theta_I|\tilde{Y}, \mathcal{M}_I) p(\Theta_I|\mathcal{M}_I) d\Theta_I \quad (44)$$

$$P(\tilde{Y}|\mathcal{M}_C) = \int \mathcal{L}(\Theta_C|\tilde{Y}, \mathcal{M}_C) p(\Theta_C|\mathcal{M}_C) d\Theta_C \quad (45)$$

where $\mathcal{L}(\cdot)$ stands for the likelihood function, and $p(\cdot|\cdot)$ denotes the posterior distribution. The model with the largest marginal data density is the one that fits the data better. We use Geweke's harmonic mean estimator (Geweke, 1999) to approximate the MDDs of these two DSGE models.

Moreover, we also consider a VAR(4):

$$\tilde{\mathbf{Y}}_t = \Phi_0 + \Phi_1 \tilde{\mathbf{Y}}_{t-1} + \Phi_2 \tilde{\mathbf{Y}}_{t-2} + \Phi_3 \tilde{\mathbf{Y}}_{t-3} + \Phi_4 \tilde{\mathbf{Y}}_{t-4} + \epsilon_t \quad (46)$$

where $\tilde{\mathbf{Y}}_t = [\ln GDP_t, \ln DEFL_t]'$ and $\Sigma_\epsilon \equiv \mathbb{E}(\epsilon_t \epsilon_t')$. We fit this VAR(4) to the same data set as that presented in section 4.1. The Minnesota random walk prior (Doan *et al.*, 1984) is implemented in order to obtain a prior distribution for the VAR parameters. Moreover, we obtain 100,000 posterior draws through Gibbs sampling. In order to compute the MDD of the VAR model we apply the method introduced by Chib (1995).

Table 3 shows that the two DSGE models are clearly misspecified, since the VAR strongly outperforms both of them in fitting the data. Nonetheless, the ICK model can be regarded as the best model in approximating the true probability distribution of the data generating process under the Kullback-Leibler distance (Fernández-Villaverde and Rubio-Ramírez, 2004).

4.5.2 IRF-based comparisons

We will assess the reliability of both the ICK model and the Calvo model in predicting how observables react to structural shocks. Since the VAR fits the data better than the two DSGE models, we can use the former as a valid benchmark to compare the IRFs of the latter. This exercise has the potential to highlight important sources of misspecification of these two DSGE models.

Let us consider the VAR(4) we introduced in the previous section. As a first step, we need to identify the shocks of this VAR. To fix notation, let us denote with Φ_ϵ the matrix such that $\epsilon_t = \Phi_\epsilon \mathbf{u}_t$, where $\mathbf{u}_t = [\varepsilon_{\Lambda,t}, \varepsilon_{a,t}]'$ is the vector of structural shocks in the DSGE models. We can decompose $\Sigma_\epsilon = \mathbf{A}\mathbf{A}'$ and introduce an orthonormal matrix $\tilde{\Omega}$, which is characterized by the rotation parameter $\tilde{\varphi} \in (-\pi, \pi]$. Hence, we can write $\Phi_\epsilon = \mathbf{A}\tilde{\Omega}(\tilde{\varphi})$. The problem of identification boils down to that of characterizing the rotation parameter $\tilde{\varphi}$.

Natural candidates of identification schemes for the VAR can be derived from the restriction (14). Nonetheless, we find that solely using this restriction delivers VAR IRFs of output to nominal shocks with implausibly large persistence. Mixing conditions derived from

the monetary policy setting of the two DSGE models and restrictions, which are consistent with other large-scale VAR studies, fixes this problem. The restrictions are presented in table 4. In this table, the monetary policy (MP) restriction is derived from the condition (14). Restriction A is consistent with the findings of Christiano *et al.* (2005), who estimate a large-scale VAR. Restriction B accords with both the ICK model and the Calvo model where real effects of monetary disturbances vanish in the long run. The purpose of the last restriction is to curb the excess persistence that would otherwise affect the VAR IRF of real output to nominal shocks. Let us express the rotation parameter that satisfies the MP restriction as $\tilde{\varphi}^*$. Moreover, for a given set of VAR parameters $(\Phi, \mathbf{A})$, the restrictions A and B characterize a set of values for the hyperparameter $\tilde{\varphi}$, which we denote as $\tilde{\Theta}$. If this set $\tilde{\Theta}$ is not empty, let us define M connected subsets $\left\{\tilde{\theta}_i\right\}_{i=1}^M$ and their lower and upper bounds $\left[\tilde{\theta}_i^L, \tilde{\theta}_i^H\right]$, such that $\tilde{\Theta} = \cup_{j=1}^M \tilde{\theta}_j$. The prior for $\tilde{\varphi}|\Phi, \mathbf{A}$ is specified as follows: if $\tilde{\Theta}$ is an empty set for given VAR parameters $(\Phi, \mathbf{A})$, $\text{prob}(\tilde{\varphi} = \tilde{\varphi}^*|\Phi, \mathbf{A}) = 1$. If $\tilde{\Theta}$ is **not** an empty set for given VAR parameters $(\Phi, \mathbf{A})$,

$$\text{prob}(\tilde{\varphi} = \tilde{\varphi}^*|\Phi, \mathbf{A}) = \frac{1}{2} \quad (47)$$

$$\text{prob}(\tilde{\varphi} \in \tilde{\Theta}|\Phi, \mathbf{A}) = \frac{1}{2} \quad (48)$$

$$\text{prob}(\tilde{\varphi} \in \tilde{\theta}_i|\Phi, \mathbf{A}, \tilde{\varphi} \in \tilde{\Theta}) = \frac{1}{M} \quad (49)$$

$$\tilde{\varphi}|\Phi, \mathbf{A}, (\tilde{\varphi} \in \tilde{\theta}_i) \sim \mathcal{U}\left[\tilde{\theta}_i^L, \tilde{\theta}_i^H\right] \quad (50)$$

where $\mathcal{U}\left[\tilde{\theta}_i^L, \tilde{\theta}_i^H\right]$ stands for the uniform distribution with mass between $\tilde{\theta}_i^L$ and $\tilde{\theta}_i^H$. Since the data are not informative about $\tilde{\varphi}$, we trivially have that

$$p\left(\tilde{\varphi}|\Phi, \mathbf{A}, \tilde{Y}\right) = p\left(\tilde{\varphi}|\Phi, \mathbf{A}\right) \quad (51)$$

As a result, the joint posterior will be

$$p\left(\tilde{\varphi}, \Phi, \mathbf{A}|\tilde{Y}\right) = p\left(\tilde{\varphi}|\Phi, \mathbf{A}, \tilde{Y}\right) \cdot p\left(\Phi, \mathbf{A}|\tilde{Y}\right) \quad (52)$$

Note that the conditional posteriors on the right-hand-side are known. Therefore, we can draw from the joint posterior by using some data-augmentation-based Monte Carlo methods.

In order to fully characterize the MP restriction, we need to set a value for λ . The posterior medians of λ implied by the ICK model and the Calvo model differ. It seems appropriate to fix $\lambda = 2$, since this value lies between the posterior medians in the two DSGE models. Nevertheless, all the results below do not significantly change by setting values for λ within an interval ranging from 1.80 and 2.16.

The IRFs of real output and inflation to a two-standard-deviation nominal shock implied by the VAR and the two DSGE models are plotted in figures 1 and 2, respectively. As also found by other studies (e.g., Christiano *et al.*, 2005), the VAR-based IRFs document highly persistent real effects of monetary disturbances. Figure 1 highlights that the Calvo model does not seem to be well-suited to accounting for such strong persistence, whereas the ICK model appears to be substantially more successful in this respect. Moreover, it is worthwhile noticing that the IRF of real output implied by the ICK model peaks three quarters after the occurrence of the shock, exactly as suggested by the benchmark VAR. On the contrary, the Calvo model predicts that the largest response of real output arises two quarters after the occurrence of the shock.

The VAR IRFs emphasize the presence of delayed effects of monetary shocks on inflation, which can be partially accommodated by the two DSGE models. Furthermore, we obtain that the IRFs of inflation implied by the two DSGE models basically overlap except at time

0. The contemporaneous response of inflation to a monetary policy shock seems to be better captured by the ICK model. Moreover, the IRF of inflation implied by the VAR reaches its peak after four quarters, while, according to the two DSGE models, this happens after three quarters.

Finally, by following Schorfheide (2008), we compute the relative reaction of inflation and output in response to a monetary disturbance implied by the ICK model, the Calvo model, and the VAR. This exercise makes the IRFs in figures 1 and 2 comparable with those implied by other DSGE models that have been estimated in the literature. We find that a 1% increase in output due to a monetary policy shock triggers an increase in the quarter-to-quarter inflation rate that ranges from 8-9 basis points for both the ICK model and the Calvo model, as well as the VAR. Schorfheide (2008) reports that a number of leading New Keynesian DSGE models predicts that this ratio ranges from 7 to 140 basis points. Thus, the degree of price flexibility predicted by the models presented in this paper is consistent with the New-Keynesian literature, even though it is relatively small.

4.6 Evaluating the predictions of the rational inattention model

In section 3.3, we showed that the rational inattention model is nested within the ICK model and sets restrictions on firms' allocation of attention (κ_m, κ_a) of the ICK model. The mapping ϕ_R in equation (41) summarizes these restrictions. Since this mapping ϕ_R is a function of only the parameters in Θ_I , we can use the posterior draws for the ICK model parameters so as to approximate the posterior distributions for the rational inattention model's predictions over κ_m^* and κ_a^* . More precisely, we implement the following procedure:

1. Sample with replacement the posterior draws we obtained when we estimated the ICK model and denote them as $\left\{ \Theta_I^{(j)} \right\}_{j=1}^M$.
2. For each sampled draw $\Theta_I^{(j)}$, compute firms' allocation of attention, $\varkappa^{(j)}, \kappa_m^{(j)}, \kappa_a^{(j)}$,

through the mapping ϕ_I in equation (39). Store the draws $\left\{\Theta_I^{(j)}, \varkappa^{(j)}\right\}_j^M$.

3. For each sampled draw $\Theta_I^{(j)}$ and associated level of attention $\varkappa^{(j)}$, use the mapping ϕ_R in equation (41) in order to get the vector (κ_m^*, κ_a^*) , that is

$$(\kappa_m^*, \kappa_a^*)^{(j)} = \phi_R \left[\tilde{\Theta}_R^{(j)}, \varkappa^{(j)} \right] \quad (53)$$

where $\tilde{\Theta}_R^{(j)}$ includes the j -th draw of parameters that belongs to $\tilde{\Theta}_R$, defined in section 3.3. Recall that $\tilde{\Theta}_R$ is a subset of Θ_I .

In practice, we set the total number of draws M equal to 1,000. In figure 3 we plot the posterior draws for parameters κ_m, κ_a implied by the ICK model (filled circles) and those for the rational inattention model's predictions κ_m^*, κ_a^* (empty squares) as well as the 45-degree line (dashed). All the plotted draws from $p(\kappa_m, \kappa_a | \tilde{Y})$ lie above the 45-degree line. This result accords well with the findings presented in section 4.4: the estimated ICK model predicts that firms allocate most of their attention to the state of technology. The rational inattention model predicts a rather balanced allocation of attention between these two shocks.

Two main factors drive the optimal allocation of attention in the rational inattention model. First, *ceteris paribus*, firms will allocate more attention to that state variable that affects more firms' expected profit function. Second, *ceteris paribus*, firms pay more attention to the state variable whose dynamics are more volatile because it is harder to keep track of it. Recall that the posterior median of the standard deviation of monetary shocks is larger than that of technology shocks (see table 2). Hence, the second effect acts to push the posterior draws for the restricted parameters κ_m^*, κ_a^* below the 45-degree line in figure 3. If the second effect were prevailing, the rational inattention model would predict that firms allocate more attention to the state of monetary policy. We do not observe such an outcome in figure 3. Therefore, the two effects act in opposite directions.

Finally, since the allocation of attention predicted by the rational inattention model is very balanced between the two state variables, we conclude that the two effects almost completely offset each other. This insight suggests that even though the predictions of the rational inattention model seem to be at odds with the data, there is room for improvement. After all, the extent to which the state of technology affects firms' expected profits is determined by only one parameter, that is, λ . Making firms' profit function less stylized has the potential to improve the fit of the model. For instance, if one allowed firms to accumulate capital, their expected profit function would be relatively more affected by technology shocks. Thus this would reinforce the first effect in a way that would push the posterior draws in figure 3 above the 45-degree line.

5 Concluding remarks

We introduce a DSGE model with imperfect common knowledge in the sense of Woodford (2002). The peculiar feature of this model is that firms do not perfectly observe the realizations of model variables. What firms observe is the history of idiosyncratic noisy signals regarding the state variables of the model, which are the aggregate technology and the monetary policy stance. Firms have to estimate the dynamics of the model variables by solving a signal extraction problem.

We fit this model to a data set that includes U.S. per capita GDP and the U.S. GDP deflator. We obtain the following results. First, when one replaces the more popular Calvo sticky pricing with the mechanism of imperfect common knowledge, the fit of the DSGE model improves. Second, we find that the mechanism of imperfect common knowledge improves upon that of sticky pricing in accounting for the persistence of real effects of monetary disturbances. Third, in the estimated model the reaction of real variables to nominal shocks is very persistent, since firms are found to be rather uninformed about the state of monetary policy.

That firms are widely unaware about monetary policy stance raises interesting questions. A natural question is: why do firms disregard the variability of the monetary policy stance, even though information about it seems to be cheaply available in advanced economies? According to the theory of rational inattention introduced by Sims (2003), free availability of a piece of information does not necessarily mean that agents will decide to pay attention to it. Hence, from a theoretical standpoint, the rational inattention theory seems to be well-suited to explaining why firms are uninformed about monetary policy even though it would be very cheap for them to become informed.

To further investigate this issue, we present a simplified rational inattention model that is nested into the model with imperfect common knowledge. Moreover, we introduce an econometric procedure that allows us to assess whether the predictions of this rational inattention model are supported by the data. We find that its predictions are rejected to some extent by the data. We point out that it is worthwhile to redo this exercise with a full-fledged model of rational inattention, where the signal process is less parametrically restricted (e.g., Maćkowiak and Wiederholt, forthcoming) or firms' profit function is less stylized (e.g., allowing firms to accumulate capital). But the lack of fast and automated routines to solve rational inattention models is a bottleneck that must be relieved in order to be able to do this exercise.

Finally, when we solve the model with information frictions, we restrict signals to be Gaussian. Sims (2006) and Lewis (2008) warn that, in models with rational inattention, such an assumption has a significant impact on agents' behavior, especially if information frictions are large. Thus, considering non-Gaussian signals is likely to affect the predictions of models with information-processing frictions. Nonetheless, some of these expansions may involve substantial technical complications. For instance, solving models with non-Gaussian signal noise may require using sequential Monte Carlo filters (Fernández-Villaverde and Rubio-Ramírez, 2007).

References

- Adam, K., 2007. Optimal monetary policy with imperfect common knowledge. *Journal of Monetary Economics* 54 (2), 267–301.
- Adam, K., 2008. Monetary policy and the great moderation, University of Mannheim Mimeo.
- An, S., Schorfheide, F., 2007. Bayesian analysis of DSGE models. *Econometric Reviews* 26 (2-4), 113–172.
- Angeletos, G.-M., La’O, J., 2008. Dispersed information over the business cycle: Optimal fiscal and monetary policy, MIT Mimeo.
- Ash, R. B., 1990. *Information Theory*. Dover Publications, New York.
- Bils, M., Klenow, P. J., 2004. Some evidence on the importance of sticky prices. *Journal of Political Economy* 112 (5), 947–985.
- Blinder, A. S., Canetti, E. R., Lebow, D. E., Rudd, J. B., 1998. *A New Approach to Understanding Price Stickiness*. Russell Sage Foundation, New York.
- Calvo, G. A., 1983. Staggered prices in a utility maximizing framework. *Journal of Monetary Economics* 12 (3), 383–398.
- Chang, Y., Doh, T., Schorfheide, F., 2007. Non-stationary hours in a DSGE model. *Journal of Money, Credit, and Banking* 39 (6), 1357–1373.
- Chib, S., 1995. Marginal likelihood from the Gibbs output. *Journal of American Statistical Association* 90 (432), 1313–1321.
- Christiano, L. J., Eichenbaum, M., Evans, C. L., 1999. Monetary policy shocks: What have we learned and to what end? In: Taylor, J. B., Woodford, M. (Eds.), *Handbook of Macroeconomics*. Vol. 1A. North Holland, New York, pp. 65–148.
- Christiano, L. J., Eichenbaum, M., Evans, C. L., 2005. Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy* 113 (1), 1–45.
- Cover, T. M., Thomas, J. A., 1991. *Elements of Information Theory*. Wiley, New York.
- Doan, T., Litterman, R., Sims, C., 1984. Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews* 3 (1), 1–100.
- Fernández-Villaverde, J., Rubio-Ramírez, J. F., 2004. Comparing dynamic equilibrium economies to data: a Bayesian approach. *Journal of Econometrics* 123 (3), 153–187.
- Fernández-Villaverde, J., Rubio-Ramírez, J. F., 2007. Estimating macroeconomic models: A likelihood approach. *Review of Economic Studies* 74 (4), 1059–1087.
- Geweke, J., 1999. Using simulation methods for Bayesian econometric models: Inference, development, and communications. *Econometric Reviews* 18 (1), 1–127.

- 654 Gorodnichenko, Y., 2008. Endogenous information, menu costs, and inflation persistence, University
655 California Berkeley Mimeo.
- 656 Hamilton, J. D., 1994. Time Series Analysis. Princeton University Press, Princeton.
- 657 Hellwig, C., 2002. Public announcements, adjustment delays and the business cycle, UCLA Mimeo.
- 658 Kass, R. E., Raftery, A. E., 1995. Bayes factors. Journal of the American Statistical Association
659 430 (90), 773–795.
- 660 Lewis, K. F., 2008. The two-period rational inattention model: Accelerations and analyses, board
661 of Governors of the Federal Reserve System: Finance and Economics Discussion Series 2008-22.
- 662 Lorenzoni, G., forthcoming A. A theory of demand shock. American Economic Review.
- 663 Lorenzoni, G., forthcoming B. Optimal monetary policy with uncertain fundamentals and dispersed
664 information. Review of Economic Studies.
- 665 Lucas, R. E. J., 1972. Expectations and the neutrality of money. Journal of Economic Theory 4 (2),
666 103–124.
- 667 Lucas, R. E. J., 1973. Some international evidence on output-inflation tradeoffs. American Economic
668 Review 63 (3), 326–334.
- 669 Luo, Y., 2008. Consumption dynamics under information processing constraints. Review of Eco-
670 nomic Dynamics 11 (2), 366–385.
- 671 Maćkowiak, B., Wiederholt, M., 2008. Business cycle dynamics under rational inattention, North-
672 western University Mimeo.
- 673 Maćkowiak, B., Wiederholt, M., forthcoming. Optimal sticky prices under rational inattention.
674 American Economic Review.
- 675 Mankiw, G. N., Reis, R., 2002a. Sticky information: A model of monetary nonneutrality and
676 structural slumps. In: Aghion, P., Frydman, R., Stiglitz, J., Woodford, M. (Eds.), Knowledge,
677 Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps.
678 Princeton University Press, Princeton, pp. 64–86.
- 679 Mankiw, G. N., Reis, R., 2002b. Sticky information versus sticky prices: A proposal to replace the
680 New Keynesian Phillips curve. Quarterly Journal of Economics 117 (4), 1295–1328.
- 681 Mankiw, G. N., Reis, R., 2006. Pervasive stickiness. American Economic Review 96 (2), 164–169.
- 682 Mankiw, G. N., Reis, R., 2007. Sticky information in general equilibrium. Journal of the European
683 Economic Association 5 (2-3), 603–613.
- 684 Nimark, K., 2008. Dynamic pricing and imperfect common knowledge. Journal of Monetary Eco-
685 nomics 55 (8), 365–382.
- 686 Paciello, L., 2008. The response of prices to technology and monetary policy shocks under rational
687 inattention, EIEF Mimeo.

- Phelps, E. S., 1970. *Microeconomic Foundations of Employment and Inflation Theory*. Macmillan, London.
- Prescott, E. C., 1986. Theory ahead of business-cycle measurement. *Federal Reserve Bank of Minneapolis Quarterly Review* 10 (Fall), 9–22.
- Reis, R., 2006a. Inattentive consumers. *Journal of Monetary Economics* 53 (8), 1761–1800.
- Reis, R., 2006b. Inattentive producers. *Review of Economic Studies* 73 (3), 793–821.
- Reis, R., 2009. A sticky-information general-equilibrium model for policy analysis, NBER Working paper 14732.
- Rondina, G., 2008. Incomplete information and informative pricing, University California San Diego Mimeo.
- Schorfheide, F., 2008. DSGE model-based estimation of the New Keynesian Phillips curve, University of Pennsylvania Mimeo.
- Shannon, C. E., 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27, 623–656.
- Sims, C. A., 2002. Solving linear rational expectations models. *Computational Economics* 20 (1-2), 1–20(21).
- Sims, C. A., 2003. Implications of rational inattention. *Journal of Monetary Economics* 50 (3), 665–690.
- Sims, C. A., 2006. Rational inattention: Beyond the linear quadratic case. *American Economic Review* 96 (2), 158–163.
- Stock, J. H., Watson, M. W., 2001. Vector autoregressions. *Journal of Economic Perspectives* 15 (4), 101–115.
- Townsend, R. M., 1983a. Equilibrium theory with learning and disparate information. In: Frydman, R., Phelps, E. S. (Eds.), *Individual Forecasting and Aggregate Outcomes*. Cambridge University Press, Cambridge, pp. 169–197.
- Townsend, R. M., 1983b. Forecasting the forecast of others. *Journal of Political Economy* 91 (4), 546–588.
- Van Nieuwerburgh, S., Veldkamp, L., 2008. Information immobility and the home bias puzzle, discussion paper, Stern School of Business, NYU.
- Woodford, M., 2002. Imperfect common knowledge and the effects of monetary policy. In: Aghion, P., Frydman, R., Stiglitz, J., Woodford, M. (Eds.), *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*. Princeton University Press, Princeton, pp. 25–58.
- Woodford, M., 2003. *Interest and Prices*. Princeton University Press, Princeton.
- Woodford, M., 2008. Information-constrained state-dependent pricing, Columbia University Mimeo.

Tables and Figures (intended for publication)

723

Table 1: Prior distributions

Name	Range	Density	Median	90% Interval
ρ_Λ	$[0, 1)$	Beta	0.50	$[0.18, 0.83]$
$\ln a$	$\mathbb{R}$	Normal	0.00	$[-0.41, 0.41]$
Λ^*	$\mathbb{R}$	Normal	0.00	$[-0.41, 0.41]$
λ	$\mathbb{R}+$	Gamma	6.67	$[0.78, 12.31]$
$100\sigma_\Lambda$	$\mathbb{R}+$	InvGamma	8.60	$[1.60, 46.60]$
$100\sigma_a$	$\mathbb{R}+$	InvGamma	0.70	$[0.51, 0.87]$
$100\sigma_{e_1}$	$\mathbb{R}+$	InvGamma	18.52	$[12.66, 25.63]$
$100\sigma_{e_2}$	$\mathbb{R}+$	InvGamma	1.00	$[0.34, 2.63]$
β	$[0, 1)$	Beta	0.99	$[0.98, 0.99]$
θ_p	$[0, 1)$	Beta	0.67	$[0.50, 0.83]$

Table 2: Posterior distributions

Name	ICK Model		Calvo Model	
	Median	95% Interval	Median	95% Interval
ρ_Λ	0.24	[0.14, 0.34]	0.06	[0.04, 0.08]
$100 \ln a$	0.45	[0.36, 0.55]	0.44	[0.29, 0.59]
$100\Lambda^*$	1.69	[1.39, 1.98]	1.85	[1.49, 2.19]
λ	1.80	[1.39, 2.21]	2.16	[1.56, 2.73]
$100\sigma_\Lambda$	1.57	[1.23, 1.92]	1.93	[1.41, 2.44]
$100\sigma_a$	0.75	[0.58, 0.90]	1.24	[0.99, 1.48]
$100\sigma_{e_1}$	36.82	[20.74, 52.21]	—	—
$100\sigma_{e_2}$	2.45	[1.37, 3.45]	—	—
β	0.99	[0.98, 0.99]	0.99	[0.99, 0.99]
θ_p	—	—	0.94	[0.92, 0.95]
$(1 - \lambda^{-1})$	0.44	[0.31, 0.56]	—	—
$\sigma_{e_1}/\sigma_\Lambda$	22.40	[15.56, 31.26]	—	—
σ_{e_2}/σ_a	3.16	[2.43, 4.04]	—	—
κ_m	0.04	[0.03, 0.06]	—	—
κ_a	0.23	[0.17, 0.29]	—	—
$\varkappa$	0.27	[0.21, 0.34]	—	—
Υ_a	0.84	[0.80, 0.89]	—	—

We use every 1,000 posterior draws to compute the posterior moments of κ_m , κ_a , $\varkappa$, and Υ_a

Table 3: Logarithms of Marginal Data Densities (MDDs)

$\log MDD$	<i>Models</i>		
	ICK	Calvo	VAR(4)
	1539.01	1530.36	1727.04

Table 4: Restrictions for identifying the VAR

MP Restriction:	$\lambda \frac{\partial GDP_t}{\partial \varepsilon_{\Lambda,t}} + \frac{\partial DEFL_t}{\partial \varepsilon_{\Lambda,t}} = 0$
Restriction A:	$\left \frac{\partial GDP_{t+j+1}}{\partial \varepsilon_{\Lambda,t}} \right - \left \frac{\partial GDP_{t+j}}{\partial \varepsilon_{\Lambda,t}} \right > 0, \text{ any } j \in \{1, 2\}$
Restriction B:	$\frac{\partial GDP_{t+50}}{\partial \varepsilon_{\Lambda,t}} < 0.005$

Restrictions refer to a two-standard-deviation monetary shock. t denotes quarters.

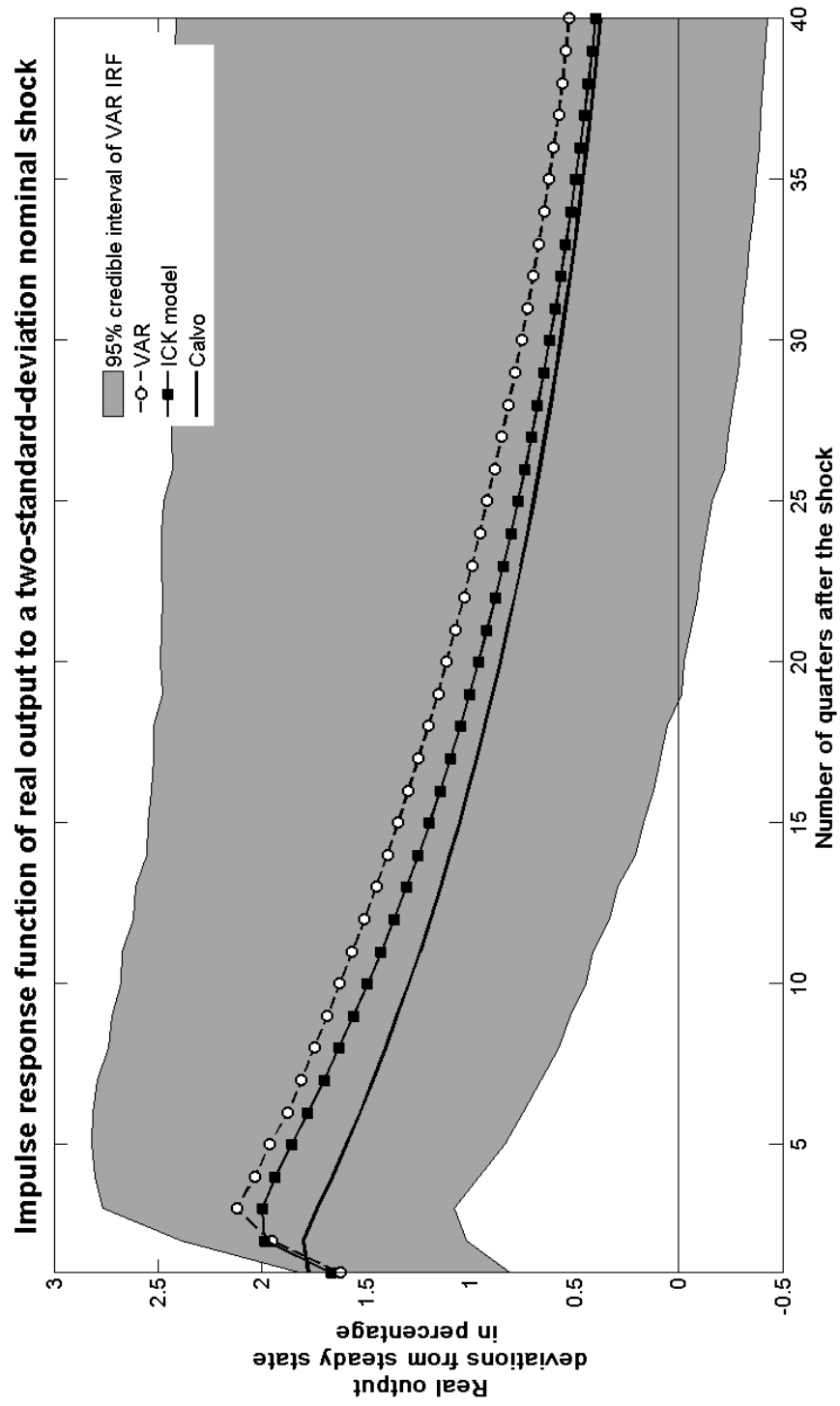


Figure 1

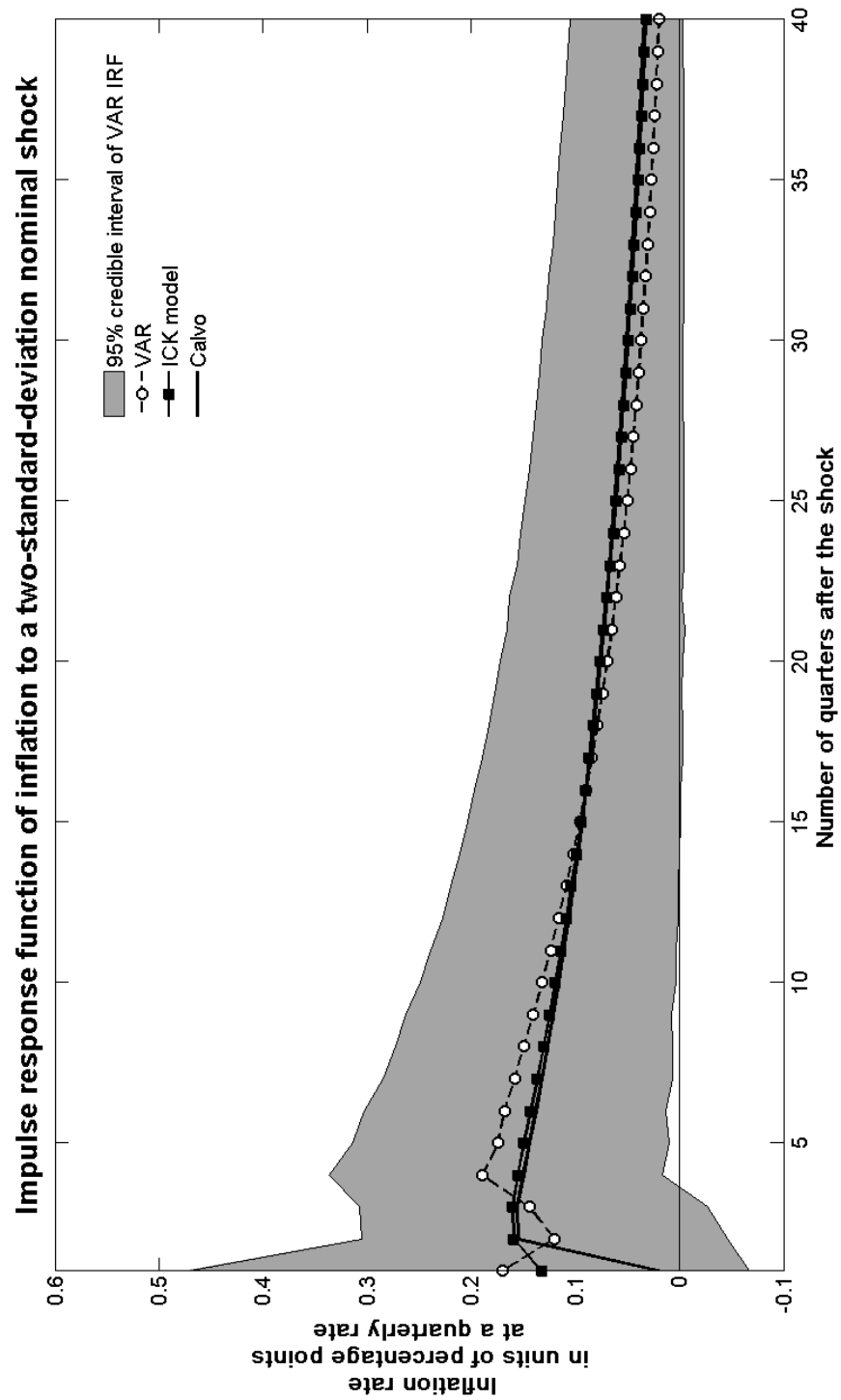


Figure 2

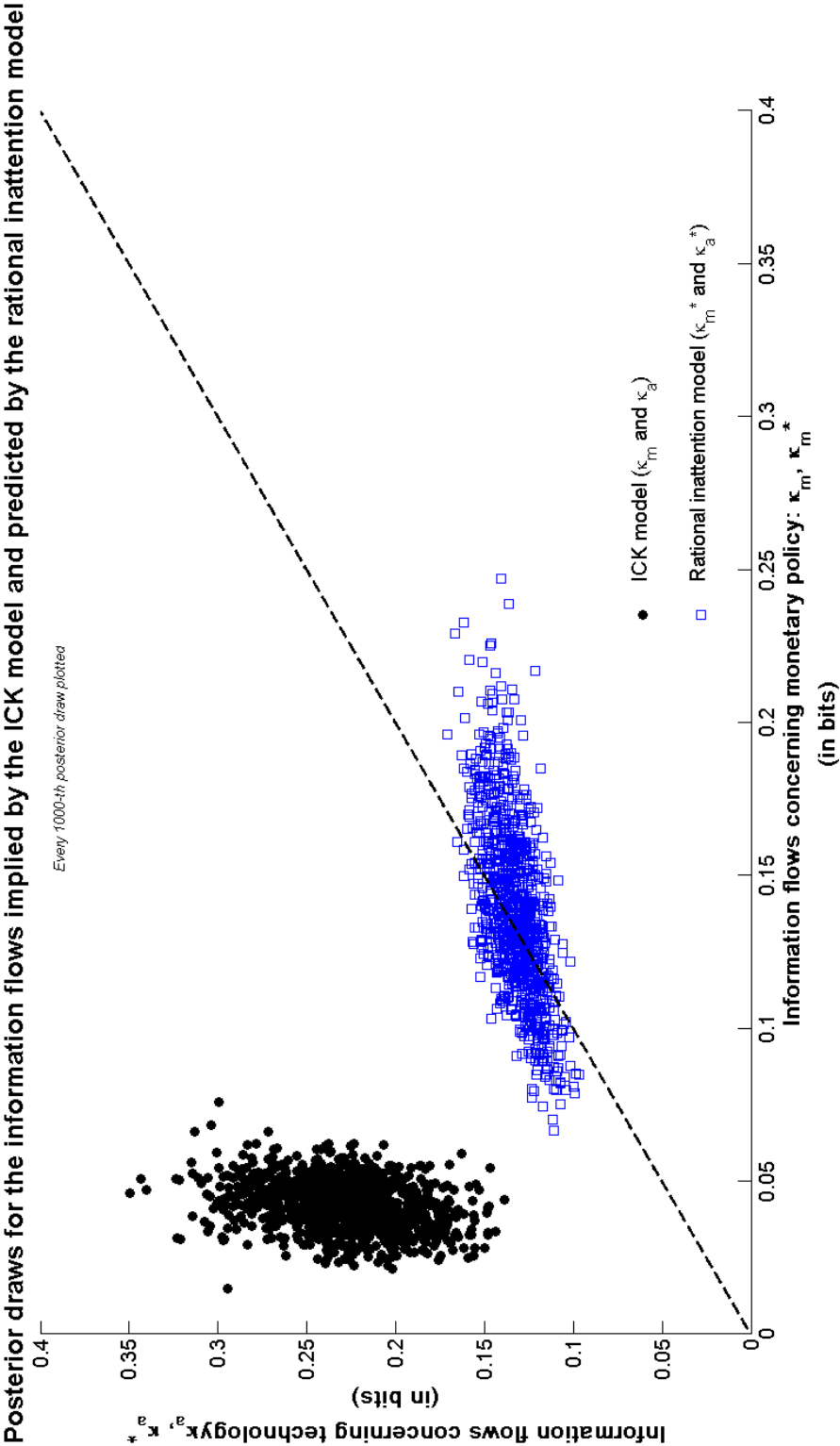


Figure 3

724 Appendix (not for publication)

725 A Log-linear approximation of the Calvo model

726 Here the equilibrium equations of the Calvo model are presented:

$$\widehat{mc}_t = \widehat{y}_t \quad (54)$$

$$\widehat{p}_t + \lambda \widehat{y}_t = 0 \quad (55)$$

$$\widehat{w}_t - \widehat{y}_t = 0 \quad (56)$$

$$\Delta a_t = \sigma_a \varepsilon_{a,t} \quad (57)$$

$$\Delta m_t = \rho_\Lambda \Delta m_{t-1} + \sigma_\Lambda \varepsilon_{\Lambda,t} \quad (58)$$

$$\widehat{\pi}_t = \beta \mathbb{E}_t \widehat{\pi}_{t+1} + \kappa_p \widehat{mc}_t - \psi_p [\lambda \Delta a_t - (1 - \Xi) \Delta m_t] \quad (59)$$

$$\widehat{\pi}_t = \widehat{p}_t - \widehat{p}_{t-1} \quad (60)$$

727 and

$$\kappa_p \equiv \frac{(1 - \theta_p \beta)(1 - \theta_p)}{\theta_p} \quad (61)$$

$$\psi_p \equiv \left(\frac{1 - 2\theta_p}{\theta_p} \right) \quad (62)$$

$$\Xi \equiv \left(\frac{1 - \theta_p}{\theta_p} + \theta_p \beta \right) \frac{\rho}{\psi_p} \quad (63)$$

728 where the subscript $\widehat{}$ means log-deviations of a variable from its flexible-price steady-state
 729 value, and mc_t denotes real marginal costs.

B Solving the models with information frictions

In section B.1, we introduce some notation in order to be able to refer to firms' higher-order beliefs. In section B.2, we outline how one can apply the method introduced by Woodford (2002) to solve the ICK model. We need to solve this model in order to evaluate the likelihood function. In section B.3, we present a method that solves the rational inattention model. We have to solve this model in order to characterize the restriction mapping ϕ_R in equation (41).

B.1 Notation of high-order beliefs

Let us consider an arbitrary stochastic variable x_t . Firm i 's expectations of order zero are the variable itself, i.e., $x_t^{(0)}(i) \equiv x_t$. Firm i 's first-order expectations are denoted as

$$x_t^{(1)}(i) \equiv \mathbb{E} [x_t | \mathcal{I}_t^i] \quad (64)$$

Average first-order expectations can be computed as follows

$$x_t^{(1)} \equiv \int x_t^{(1)}(i) di \quad (65)$$

Firm i 's second-order expectations are firm i 's first-order expectations of the average first-order expectations, or more concisely

$$x_t^{(2)}(i) \equiv \mathbb{E} [x_t^{(1)} | \mathcal{I}_t^i] \quad (66)$$

By rolling this argument forward we obtain the average m -th order expectation,

$$x_t^{(m)} \equiv \int x_t^{(m)}(i) di \quad (67)$$

Moreover, firm i 's $(m+1)$ -th order expectations are its expectations of the average m -th order expectation,

$$x_t^{(m+1)}(i) \equiv x_{t|t}^{(m)}(i) \equiv \mathbb{E} \left[x_t^{(m)} | I_t^i \right] \quad (68)$$

746 Firms' price-setting equation can be shown to be:

$$\ln P_t^i = (1 - \lambda^{-1}) \ln P_t^{(1)}(i) + \lambda^{-1} \ln \Lambda_t^{(1)}(i) - \ln A_t^{(1)}(i) - \ln \bar{y} \quad (69)$$

747 where $\ln P_t^{(1)}(i)$, $\ln \Lambda_t^{(1)}(i)$, and $\ln A_t^{(1)}(i)$ stand for firm i 's first-order expectations of $\ln P_t$,
748 $\ln \Lambda_t$, and $\ln A_t$, respectively.

749 By aggregating across firms we obtain the price equation

$$\ln P_t = (1 - \lambda^{-1}) \ln P_t^{(1)} + \lambda^{-1} \ln \Lambda_t^{(1)} - \ln A_t^{(1)} - \ln \bar{y} \quad (70)$$

750 where $\ln P_t^{(1)}$, $\ln \Lambda_t^{(1)}$, and $\ln A_t^{(1)}$ are the average first-order expectations of $\ln P_t$, $\ln \Lambda_t$, and
751 $\ln A_t$, respectively.

752 Iterating on equation (70) by repeatedly taking conditional expectations and averaging
753 across firms yields the law of motion of the price level:

$$\ln P_t = \left[\sum_{j=1}^{\infty} (1 - \lambda^{-1})^{j-1} \lambda^{-1} \left(\ln \Lambda_t^{(j)} - \lambda \ln A_t^{(j)} \right) \right] - \lambda \ln \bar{y} \quad (71)$$

754 B.2 Solving the ICK model

755 For a given set of parameters Θ_I , the transition equations of the ICK model are:

$$\hat{y}_t = -\lambda^{-1} \hat{p}_t \quad (72)$$

$$\hat{p}_t = \mathbf{r}' \bar{\mathbf{X}}_t \quad (73)$$

$$\bar{\mathbf{X}}_t = \bar{\mathbf{B}} \bar{\mathbf{X}}_{t-1} + \bar{\mathbf{b}} \mathbf{u}_t \quad (74)$$

756 where

$$\bar{\mathbf{X}}_t \equiv \left[\mathbf{X}'_t : \mathbf{F}'_t \right]', \mathbf{r} \equiv [-1, 0, 1, 1, 0, -1]' \quad (75)$$

$$\mathbf{F}_t \equiv \sum_{j=1}^{\infty} (1 - \lambda^{-1})^{j-1} \lambda^{-1} \mathbf{X}_t^{(j)} \quad (76)$$

$$\mathbf{X}_t \equiv [m_t, m_{t-1}, \lambda a_t] \quad (77)$$

$$\bar{\mathbf{B}} \equiv \begin{bmatrix} \mathbf{B}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{G}_{3 \times 3} & \mathbf{H}_{3 \times 3} \end{bmatrix}, \quad \bar{\mathbf{b}} = \left[\mathbf{b}' : \mathbf{d}' \right]' \quad (78)$$

$$\mathbf{B} \equiv \begin{bmatrix} 1 + \rho_{\Lambda} & -\rho_{\Lambda} & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} \equiv \begin{bmatrix} \sigma_{\Lambda} & 0 \\ 0 & 0 \\ 0 & \lambda \sigma_a \end{bmatrix}, \quad \mathbf{u}_t = [\varepsilon_{\Lambda, t}, \varepsilon_{a, t}]' \quad (79)$$

$$\mathbf{u}_t \stackrel{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_2), \text{ for all } t \quad (80)$$

757 where $\mathbf{I}_2$ is a 2×2 identity matrix,

$$\mathbf{G} = \tilde{\mathbf{k}} \mathbf{B}^{\dagger}, \mathbf{d} = \tilde{\mathbf{k}} \Sigma^{\frac{1}{2}}, \mathbf{H} = \mathbf{B} - \tilde{\mathbf{k}} \mathbf{B}^{\dagger} \quad (81)$$

$$\tilde{\mathbf{k}} \equiv \boldsymbol{\varphi}' \mathbf{k}, \boldsymbol{\varphi} \equiv \left[\lambda^{-1} \cdot \mathbf{I}_3 : (1 - \lambda^{-1}) \cdot \mathbf{I}_3 \right]' \quad (82)$$

$$\Sigma = \begin{bmatrix} \sigma_{\Lambda} & 0 \\ 0 & \sigma_a \end{bmatrix} \quad (83)$$

$$\mathbf{D} \equiv \left[\mathbf{D}_1 : \mathbf{0}_{2 \times 3} \right], \quad \mathbf{D}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1/\lambda \end{bmatrix} \quad (84)$$

758 where $\mathbf{B}^{\dagger} \equiv \left[\mathbf{B}'_1 \quad \frac{1}{\lambda} \mathbf{B}'_3 \right]'$ and $\mathbf{B}_j$ stands for the j -th row of $\mathbf{B}$ and $\mathbf{k}$ is the steady-state
759 matrix of Kalman gains which is well-known to be equal to

$$\mathbf{k} = \mathbf{P} \mathbf{D}' [\mathbf{D} \mathbf{P} \mathbf{D}' + \Sigma_e]^{-1} \quad (85)$$

with Σ_e defined as in (21) for the ICK model. The variance and covariance matrix $\mathbf{P}$ solves the following algebraic Riccati equation:

$$\mathbf{P} = \overline{\mathbf{B}} \left[\mathbf{P} - \mathbf{P} \mathbf{D}' [\mathbf{D} \mathbf{P} \mathbf{D}' + \Sigma_e]^{-1} \mathbf{D} \mathbf{P} \right] \overline{\mathbf{B}}' + \overline{\mathbf{b}} \overline{\mathbf{b}}' \quad (86)$$

A loop to numerically find an REE for the ICK model is as follows. Given a set of parameter values and a guess for the Kalman-gain matrix $\mathbf{k}^0$, one has to characterize the matrices $\mathbf{G}$, $\mathbf{H}$, and $\mathbf{d}$. Then, one has to solve the algebraic Riccati equation (86) for $\mathbf{P}$ and to obtain a new Kalman-gain matrix $\mathbf{k}^*$ through equation (85). Then if the new Kalman-gain matrix is sufficiently close to the guess, one has just found the fixed point and stops; otherwise, one goes through another loop by using the matrix $\mathbf{k}^*$ as a new guess for the Kalman-gain matrix. Once a fixed point is found, one can use the resulting Kalman-gain matrix in order to fully characterize the state-space system described in (72)-(84). Computationally, finding this fixed point turns out to be very fast and this makes the ICK model suitable for estimation.

B.3 Solving the rational inattention model

The following algorithm allows one to solve the rational inattention model. Given a set of parameter values for Θ_R ,

1. **GUESS:** Guess σ_{e_1} and σ_{e_2} .
2. **DETERMINING THE EQUILIBRIUM TRANSITION EQUATION OF FIRMS' STATE SPACE MODEL:** Use the parameter values:

$$\{\nu, \rho_\Lambda, \alpha, \ln a, \Lambda^*, \lambda, \beta, \sigma_\Lambda, \sigma_a\} \subset \Theta_R$$

as well as the guessed parameters σ_{e_1} and σ_{e_2} to apply the method shown in Appendix A1. Store the variance-covariance matrix $\mathbf{P}$, defined in equation (86).

3. **SOLVE FIRMS' ATTENTION PROBLEM:** Solve the quadratic approximation of firms' attention problem:

$$(\sigma_{e_1}^*, \sigma_{e_2}^*) = \arg \min_{st} VAR(\ln P_{i,t}^\diamond | \mathbf{z}_i^t) \quad (87)$$

$$H(m_t | z_{1,i}^{t-1}) - H(m_t | z_{1,i}^t) + H(a_t | z_{2,i}^{t-1}) - H(a_t | z_{2,i}^t) \leq \kappa \quad (88)$$

where $\ln P_{i,t}^\diamond$ is the log of optimal price set by firm i at time t under perfect information (i.e., $\kappa \rightarrow \infty$), which can be shown to be:

$$\ln P_{i,t}^\diamond = (1 - \lambda^{-1}) \ln P_t + \lambda^{-1} \ln \Lambda_t - \ln A_t + \lambda^{-1} \ln \bar{p} \quad (89)$$

Notice that one can rewrite the objective function as

$$\begin{aligned} VAR(\ln P_{i,t}^\diamond | \mathbf{z}_i^t) &= (1 - \lambda^{-1})^2 VAR(\ln P_t | \mathbf{z}_i^t) + \lambda^{-2} VAR(m_t | \mathbf{z}_i^t) + VAR(a_t | \mathbf{z}_i^t) \\ &\quad + 2\lambda^{-1} (1 - \lambda^{-1}) cov(m_t \ln P_t | \mathbf{z}_i^t) - 2(1 - \lambda^{-1}) cov(\ln P_t a_t | \mathbf{z}_i^t) \end{aligned} \quad (90)$$

All these conditional variances and covariances are obtained from the matrix $\mathbf{P}$ that we characterized at step 2. Note that the equilibrium law of motion of the price level (71) implies that

$$VAR(\ln P_t | \mathbf{z}_i^t) = \mathbf{u}' \mathbb{E} [\bar{\mathbf{X}}_t \bar{\mathbf{X}}_t'] \mathbf{u} \quad (91)$$

with $\mathbf{u} \equiv [0, 0, 0, 1, 0, -1]'$. Moreover, denote the (i, j) element of the matrix $\mathbf{P}$ as $\mathbf{P}(i, j)$. It is easy to see that

$$VAR(m_t | \mathbf{z}_i^t) = \mathbf{P}(1, 1); \quad VAR(a_t | \mathbf{z}_i^t) = \mathbf{P}(3, 3)$$

$$cov(m_t \ln P_t | \mathbf{z}_i^t) = \mathbf{P}(4, 1) - \mathbf{P}(6, 1);$$

$$cov(\ln P_t a_t | \mathbf{z}_i^t) = \mathbf{P}(4, 3) - \mathbf{P}(6, 3)$$

Finally, from equations (30)-(33), we observe that the information-processing constraint (88) can also be numerically characterized by using the matrix $\mathbf{P}$ from step 2.

4. **CHECK THE GUESS:** Check if $\|\sigma_{e_j}^* - \sigma_{e_j}\| \leq \varepsilon$, with $j \in \{1, 2\}$, and $\varepsilon > 0$ small. If this criterion is not satisfied, go back to step 1 by setting $\sigma_{e_j} = \sigma_{e_j}^*$, $j \in \{1, 2\}$. Otherwise STOP.