

Monetary Shocks in Models with Inattentive Producers

FERNANDO E. ALVAREZ

University of Chicago and NBER

FRANCESCO LIPPI

Einaudi Institute for Economics and Finance, University of Sassari, and CEPR

and

LUIGI PACIELLO

Einaudi Institute for Economics and Finance and CEPR

First version received November 2012; final version accepted August 2015 (Eds.)

We study models where prices respond slowly to shocks because firms are rationally inattentive. Producers must pay a cost to observe the determinants of the current profit maximizing price, and hence observe them infrequently. To generate large real effects of monetary shocks in such a model the time between observations must be long and/or highly volatile. Previous work on rational inattentiveness has allowed for observation intervals that are either constant-but-long (*e.g.* Caballero, 1989 or Reis, 2006) or volatile-but-short (*e.g.* Reis's, 2006 example where observation costs are negligible), but not both. In these models, the real effects of monetary policy are small for realistic values of the duration between observations. We show that non-negligible observation costs produce both of these effects: intervals between observations are infrequent *and* volatile. This generates large real effects of monetary policy for realistic values of the average time between observations.

Key words: Observation costs, Inattentiveness, Monetary shocks, Impulse responses

JEL Codes: E5

1. INTRODUCTION

The sluggish propagation of new information from individual into aggregate prices is one of the key mechanisms behind the transmission of monetary shocks and the analysis of their real effects. We study microfounded models where prices respond slowly to shocks because firms must pay a fixed cost to observe the determinants of the profit maximizing price, as in the “rational inattentiveness” literature. The timing of these observations is random, and the firm’s frequency of observation determines the velocity with which new information is embedded into prices, so that large real effects of monetary shocks require infrequent observations. For a given mean frequency of observation, a larger volatility of the time between observations increases the real effects, a result established by Carvalho and Schwartzman (2012) for the case of i.i.d. observation times. The models in Caballero (1989), and the baseline model of Reis (2006), yield optimally chosen constant observation times that produce infrequent adjustments but have no volatility. These models yield real effects similar to Taylor’s exogenously staggered adjustment model, and about half the size of what is produced by the Calvo model or the exponentially

distributed observations as used in Mankiw and Reis (2002), parametrized to have the same mean frequency of observations as the model with constant times. For a special case with negligibly small observation costs Reis (2006) provides a microfoundation for random observation times.¹ Although the times between observations are volatile in this case, the frequency of observations is very high (due to the negligible costs), so that the real effects of monetary shocks are tiny.

Motivated to obtain larger effects of monetary shocks, we study models where the intervals between optimally chosen observation times are both *long* and *volatile*. The first set of results derives the firms' observation times as the solution to a fully specified profit maximization problem, with random and persistent variation in the observation costs. To obtain long and volatile observation times, and hence larger effects of monetary policy, it is necessary to have observation costs that are both non-negligible and sufficiently persistent. Indeed, an important innovation of this article is to produce from first principles optimal firm decisions that are persistent rather than i.i.d., as is standard in the previous literature. We show that the process of the observation costs maps into the process for the optimally chosen observation times in a subtle non-linear way. For instance, a setting with i.i.d. observation costs will produce optimally chosen *constant* observation times. As the costs become sufficiently persistent, the optimal times become responsive to the realization of the observation cost, and hence the frequency of observation becomes volatile and persistent. The second set of results pertains to the aggregation properties of an economy with a continuum of firms, for a given Markov process for the individual firm's observation times. Methodologically, this aggregation result is new and generalizes the elementary renewal theorem in statistics where consecutive holding times (*i.e.* times between observations in our economic application) are i.i.d. to an environment where holding times follow a first-order Markov process. The paper contains several new results, summarized next, that extend and complement the ones in the literature.

Microfoundations of optimal observation times: We use a simple general equilibrium model of price setting to derive the distribution of the optimal times between observations from the firm's profit maximization problem. We assume that when the firm pays the observation cost it learns the current value of the production cost, a key piece of information for price setting, and also obtains a signal about future observation costs. The optimal choice for the time until the next observation is a function of the signal, a sufficient statistic for the future observation cost at every future date. We consider several cases, which are interesting both for their relationship to the literature, and for their substantive monetary policy implications.

The first case is one where the signals are such that the optimally chosen observation times are i.i.d. This case, while standard in the "rational inattentiveness" literature, has not yet been derived as the solution of an explicit profit maximization problem with random variation in the benefit/cost of acquiring information.² This case yields two novel insights. First, i.i.d. observation

1. The case of negligibly small observation cost studied by Reis (2006) is one where for any given distribution of the observation cost θ , the model is solved for a scaled process of the observation cost $\theta' = c\theta$ where $c \rightarrow 0$ is a scaling factor. In the scaled model, both the mean and volatility of the observation cost go to zero, but the coefficient of variation stays constant.

2. An exception is the paper by Woodford (2009) who uses a different framework, namely the one of "rational inattention", to characterize optimal i.i.d. observation times. He assumes that keeping track of the state and *time elapsed* is costly. When the cost of keeping track of time is sufficiently large the optimal observation times become i.i.d. because the firm cannot base decisions on (any measure of) time and, therefore, base decisions exclusively on the current value of the noisy signals it receives, even though the process of interest is persistent. Each of the two frameworks has weaknesses and strengths in this particular application. On the one hand, we view the assumption of no memory using the rational inattention model of Woodford as an extreme one. On the other hand, in the rational inattention model the design of the signals about the state space is chosen by the decision maker. Instead, in the rational inattentiveness model that we use, the signals are extreme: either complete or no information is revealed upon observation.

times arise only if the evolution of the observation costs is not independent of the firm's decision to observe. Thus, the firm gets a new cost every time it makes an observation. This requirement is specific to the i.i.d. case, and illustrates how special this arrangement is. The more natural setting in which the evolution of the costs is independent of the observation times will naturally give rise to Markov observation times, discussed next. The second insight yielded by the model with i.i.d. observation times is that the optimal decision by the firm contains an option value argument: a firm that is facing a very low cost (for the next observation), has an incentive not to observe immediately, but rather to save the opportunity for a cheap observation for when it is most valuable in future. This implies that the support of the distribution of observation cost must include negative values to generate the arbitrarily small observation times that are common in the literature, such as the exponentially and i.i.d. distributed observations in Mankiw and Reis (2002). Both insights suggest that the i.i.d. case is rather special, for this reason we explore the more general, and more realistic case, which gives rise to persistent observation times.

The second case we explore assumes that the evolution of the costs is independent of the firm's decision to observe. In this case the optimal decision rules imply that the times between successive observations form a Markov process. We characterize the properties of the observation cost process that leads to variability in the times between observations. This is important since, as mentioned, the variability of the times between observations affects the real output effect of an aggregate demand shock. Notice that variable observation times require both variability in costs and that the optimal decision rule is sensitive to the realization of the cost shock. An early contribution for volatile observation times was given by Reis (2006), who provides a microfoundation in the limiting case of negligible observation costs. His model makes the times between observations random, and hence volatile, but it produces very frequent observations (due to the small costs), so that the real effects of monetary shocks are tiny. We provide a global characterization of the decision rule, which allows for both negligible (reproducing Reis' result) as well as non-negligible observation costs. The latter are essential to get infrequent observations. Interestingly, we show that for any given variability of the observation costs, the *persistence* of the costs affects the sensitivity of the optimal decision rule with respect to the cost. Hence, the persistence of the observation costs determines the variability of times between observations. Higher persistence amplifies the variability of optimal observation times and, therefore, the real effects of monetary shocks. On the one hand, we show that for any given (non-negligible) expected value of the observation costs, the volatility of observation times converges to zero if persistence is sufficiently low. In the limiting case where the observation costs are i.i.d., the volatility of times between observations converges to zero and the real effects of monetary shocks are those of a model with constant times between observations. On the other hand, the real effects of monetary shocks can be substantially larger, close to those of a Calvo model with a realistic parametrization, with observation costs that are non-negligible and sufficiently persistent.

The analysis in this article will mostly focus on a model where firms are ex ante identical but differ ex post in their choices about the optimal observation times, which follow a Markovian process. There are two reasons for this choice. First, as argued above, this case emerges naturally from first principles in models where the opportunity cost of an observation is persistent. Secondly, we notice that an alternative model with constant observation costs that differ across firms might also be able to produce larger output effects. Such a model, however, implies a constant time between observations, say T_i , for a given firm i , *i.e.* a hazard of price adjustment equal to zero and with a vertical asymptote at T_i . Recent empirical research on the behaviour of individual prices provides ample evidence that there is large variation in the duration between price changes *at the UPC-store level*, *e.g.* Alvarez et al. (2015). Given the usefulness of mapping the predictions of the model to available micro data (*e.g.* for the hazard rate and the distribution of the size of price changes), we see the characterization of the Markov case as more valuable.

Results on aggregation: Another set of results pertains to the aggregation properties of an economy with “inattentive producers”. We denote by $H(t'|t)$ the right CDF of the optimally chosen times larger than t' , conditional on a gap of time equal to t between the preceding two observations. This distribution is derived simply from the assumptions about the observation costs, the signals and the optimal decision rules described above. While in our model the distribution $H(t'|t)$ is the outcome of the solution to the firms’ problem facing random and persistent variability in observation costs, our results on aggregation are independent of the specific model micro-founding a given H . In particular, for a given H , we aggregate the times until the next observation across a continuum of firms to characterize the stationary cross-sectional distribution of the times until the next observation, which we denote by Q . The distribution Q is key in determining the propagation of monetary shocks as it determines the time it takes before a given fraction of firms makes at least one observation and, therefore, incorporates new information into prices.³ We show that the density q associated with the cross-sectional distribution Q is given by $q(t) = RK(t)$, where R is a constant (the unconditional frequency of observations per unit of time), and K is the invariant distribution of the Markov process formed by the time elapsed between consecutive observations.⁴ And so by the law of large numbers, $K(t)$ measures the cross-sectional fraction of firms that have waited at least t periods since the last observation.

We use this result to illustrate several applications. First, we show that the cumulative output response increases with both the *average* and the *coefficient of variation* of times between consecutive observations, as measured in the cross-section according to the distribution K . This finding generalizes a result by Carvalho and Schwartzman (2012), obtained for the case of i.i.d. durations, to a framework where durations form a first-order Markov process. This extension is important because quite special assumptions are needed to generate i.i.d. durations, as discussed above. This result is quantitatively relevant because persistent shocks, and the associated persistent times between observations, are essential to generate a sizeable cross-sectional variation of observation times from a reasonable variation in the observation cost.

Secondly, our analysis of the mapping from H to Q clarifies a related theoretical result in Reis (2006) concerning the aggregation of individual firms decision. Within the class of i.i.d. observation times, Reis derives conditions on the individual firm’s distribution of observation times that deliver an exponential cross-sectional distribution Q , as assumed *e.g.* in Mankiw and Reis (2002). We clarify that a necessary and sufficient condition to obtain an exponential cross-sectional distribution Q is that the firm-level distribution of observation times H must be exponential itself. Hence, the exponential distribution of observation times, a case often used in the literature, turns out to be obtained as an optimal decision rule only in a very special case.⁵

Thirdly, under the assumption that price changes only occur upon observation dates, we derive a mapping between the cross-sectional distribution of observation durations, K , and the kurtosis of the size-distribution of price changes.⁶ This mapping can be used to calibrate the model.

3. Reis (2006) labels Q the “distribution of inattentiveness”.

4. Formally, $K(t)$ is the invariant cumulative distribution associated with the transition function $H(t'|t)$, as standard in the definition of the invariant distributions for a discrete-time continuous-state Markov process.

5. Indeed, we solve analytically for the distribution of costs that implies an exponential distribution of observation times, which has two features that we find implausible: a large proportion of costs that are *negative* and a very large variability of observation costs.

6. A microfoundation of price changes occurring only upon an observation dates can be the presence of menu costs. As shown in Proposition 1 of Alvarez *et al.* (2011), relatively small menu costs make price plans not optimal in this environment at moderate inflation rates.

Alternative microfoundations of random observation times: The aggregation results and the characterization of the cumulative IRF of a monetary shock in terms of distribution of observation times hold for a large class of models, beyond those with random observation *costs*. For instance, the results hold in models where the *benefits* of observation are random. In the Online Appendix we give an example of such a model in which the observation costs are constant but the benefits of observing are random due to a time-varying volatility of the production cost.

The rest of the article is organized as follows. For expositional convenience, we first derive the mapping from the individual firm's distribution of the times between observations H to the cross-sectional distribution Q (Section 2), and discuss three main applications of this new aggregation result. Section 3 describes a general profit maximization problem, in the presence of a fixed cost of observing the state. The solution to this problem yields the optimal decision rules that micro-found the stochastic observation times of an individual firm. Section 4 shows the general solution to the firm's problem, in which case the optimal observation times form a first-order Markov process. Section 5 consider a particular setup for the firm's problem, in which case the optimal observation times are i.i.d. Section 6 concludes.

2. AGGREGATING THE BEHAVIOUR OF INATTENTIVE PRODUCERS

This section characterizes the linkages between an individual firm's behaviour and the cross-sectional features that are important for modelling the propagation of an aggregate shock in an economy with "rationally inattentive" firms. We are interested in an economy populated by firms that observe the state underlying their prices at randomly spaced times. The interpretation of "making an observation" is that the firm pays a cost to gather (aggregate and process) information to be used to set prices. Upon each observation the firm chooses a new path for prices until the next observation.

The primitive for the analysis in this section is a distribution of the firm's times between consecutive observations, as summarized by its right cumulative distribution function (CDF), which we denote by H .⁷ Starting from a given H , this section aggregates the times between observations across firms to characterize the stationary *cross-sectional* distribution of the "times until the next observation". That is, the fraction of firms that, at any point in time, will wait at least t units of time until the next observation. We denote the right CDF of such distribution by $Q(t)$.

The distribution $Q(t)$ is a key object in models with rational inattentive producers because it determines the time left before a given fraction of firms will conduct an observation and adjust their behaviour to the new information obtained. Therefore, Q directly determines the time it takes for an aggregate shock to be incorporated into the information set of a given fraction of firms. Thus, Q determines the speed at which a monetary shock affects the aggregate price level. In a stationary environment there is a constant rate of observations per unit of time, which we denote by R . In this section, we will derive both Q and R as implied by a given H . In Section 3, we will study a model where the distribution H is itself derived from the firm's optimal price setting choices subject to an information-gathering friction.

Before providing an exhaustive formal analysis, we sketch two examples to fix ideas. The first one is a classic Taylor's model where each firm observes the state at deterministically spaced time intervals of length T . This assumption amounts to an i.i.d. $H(t)$ distribution which is piecewise continuous: constant and equal to 1 for $t \in [0, T)$, and zero otherwise. It is easy to see how this $H(t)$

7. The use of right CDF, *i.e.* the probability mass that the time between observations is *above* a given threshold, is more convenient algebraically than the traditional (left) CDF.

implies a unique invariant cross-sectional distribution $Q(t)$ of times until the next observation, with a density $q(t)$ that is uniform for $t \in (0, T)$, and an observation rate given by $R = 1/T$. The second example is the widely used Calvo pricing. It assumes that the distribution of times between observations for a firm follows a Poisson process, so that the probability that a firm waits at least t periods after adjusting is given by the exponential distribution $H(t) = e^{-\lambda t}$ for $t \in (0, \infty)$. Aggregating a continuum of firms that follow this $H(t)$ rule gives a cross-sectional distribution of times until the next observation, $Q(t)$, that is also exponential, and a rate of observations per unit of time $R = \lambda$.

While the mapping from H to Q for some simple examples can be solved using intuition, a general treatment is useful. It will allow us to extend the mapping along two important dimensions: first, to cases where the firm hazard rate of observations is not constant (so that $H(t)$ is not exponential); secondly, and perhaps more interestingly, to allow for non-i.i.d. observation times at the firm level. In the latter, the distribution of the times t between observations is allowed to depend on the length s of the time since the last observation and denoted by $H(t|s)$. Markovian observation times appear to be a robust feature of a microfounded model, as we show in Section 3, we thus see this case as an important extension of the i.i.d. observation times.⁸

We consider setups where the optimal decision of the firm will be such that, in a stationary environment, the time elapsed between consecutive observations forms a stationary Markov process. As already mentioned, the primitive of the analysis in this section is the *conditional* right CDF $H(t|s)$ for the times between consecutive observations of a firm. The function $H: \mathbb{R}_+^2 \rightarrow [0, 1]$ satisfies the following: (1) $H(0|s) = 1$ for all $s \geq 0$, (2) $H(t|s) \geq H(t'|s)$ for all $t' \geq t$, (iii) $\lim_{t \rightarrow \infty} H(t|s) = 0$ for all $s \geq 0$, and iv) $H(t|\cdot)$ is Borel-measurable, for all $t \geq 0$. The interpretation of $H(t|s)$ is that if the gap between the preceding two observations is equal to s , then the probability that the time between the current and the next observation is at least t is $H(t|s)$. We assume that the stochastic process for observation times is independent across firms, and that the law of large number holds.

Let $K(s)$ denote the invariant cumulative distribution function of the times between consecutive observations of length up to s . Formally, $K(s)$ is the invariant distribution associated with the transition function $H(t'|t)$, as will be shown below. In particular, in an interval of time $(t, t+dt)$ there are Rdt observations that take place. Of those, a fraction $K(s)$ is characterized by a gap between the preceding two observations greater or equal than s . Using the definition of $K(s)$ we can compute the cross-sectional distribution of times until the next observation, whose right CDF is⁹

$$Q(t) = -R \int_0^\infty \left(\int_0^\infty H(u+t|s) dK(s) \right) du = -R \int_t^\infty \int_0^\infty H(u|s) dK(s) du. \quad (1)$$

The expression is easy to interpret: the firms that will wait at least t periods before the next observation are those that drew an observation time larger than $u+t$ some u periods ago. The integral inside the parenthesis measures those firms. The probability that a firm must wait this long depends on the gap s between the preceding two observations and is given by $H(u+t, s)$. In the integrand, this is then weighted by (the negative of) $dK(s)$, which gives the fraction of firms whose preceding two observations were separated by s . Integrating over both the time until the next observation, u , and the time between the preceding two observations, s , and multiplying by the frequency of observation, R , gives $Q(t)$. Note that the minus sign comes from using a *right* CDF for $K(s)$. The value of R is easily computed from equation (1), using that $Q(0) = 1$.

8. See Alvarez *et al.* (2011) for a microfounded model where the optimal observation times are not i.i.d. We note that there is no systematic treatment of this problem in the literature.

9. Integration follows Lebesgue–Stieltjes. $Q(t)$ is also referred to as distribution of the forward times.

Differentiating this expression we obtain the density of the cross-sectional distribution of times until the next observation, as we summarize in the following lemma:

Lemma 1. *Given the conditional right CDF, H , and its corresponding invariant distribution of times between observations, K , the density of the times until the next observation $q(t)$ is:*

$$q(t) \equiv -\frac{\partial Q(t)}{\partial t} = -R \int_0^\infty H(t|s) dK(s) \geq 0, \tag{2}$$

where the number of observations per unit of time is

$$R = -\frac{1}{\int_0^\infty \int_0^\infty H(u|s) dK(s) du} \geq 0. \tag{3}$$

To complete the characterization of Q we need the (invariant) distribution of times between observations, $K(s)$. We distinguish between two cases: one where the distribution H is absolutely continuous, another where H is discrete. If $H(\cdot|s)$ is absolutely continuous the invariant distribution K will be differentiable with density $k(t) = -\partial K(t)/\partial t$ solving:

$$k(t) = \int_0^\infty h(t|s) k(s) ds \text{ for all } t \geq 0 \text{ where } h(t|s) \equiv -\frac{\partial H(t|s)}{\partial t} \geq 0 \text{ for all } t, s \geq 0.$$

Likewise, since H is differentiable almost everywhere, the right CDF of the invariant measure solves

$$K(t) = -\int_0^\infty H(t|s) dK(s) \text{ for all } t \geq 0. \tag{4}$$

The second case occurs when $H(\cdot|s)$ is a step function, so that there are countably many observation times \mathcal{T} with non-zero probabilities. Let $\mathcal{T} = \{t_1, \dots, t_i, \dots, t_I\}$ with $t_i < t_{i+1}$ for $i \geq 1$ and $h = [h_{t_i, t_j}]$ be a matrix defined as

$$h_{t_i, t_j} \equiv \lim_{t \uparrow t_i} H(t|t_j) - H(t_i|t_j) \geq 0.$$

Using this notation, $\{k(t)\}_{t \in \mathcal{T}}$ is the corresponding invariant measure, *i.e.* a positive eigenvector associated with h solving:

$$k(t_i) = \sum_{t_j \in \mathcal{T}} h_{t_i, t_j} k(t_j) \text{ for all } t_i \in \mathcal{T}. \tag{5}$$

In this case, the invariant measure is given by:

$$K(t) = \begin{cases} 1 & \text{if } t < t_1 \\ \sum_{t_i > t} k(t_i) & \text{if } t_1 \leq t_i < t_I, \\ 0 & \text{if } t \geq t_I \end{cases} \tag{6}$$

where the $k(t_i)$ elements solve equation (5). We collect these results in the following lemma:

Lemma 2. *Given a conditional right CDF H for the times between consecutive observations, the invariant distribution K is given by the solution of equation (4) if $H(\cdot|s)$ is absolutely continuous for each t , or by the solution of equation (6) if $H(\cdot|s)$ is a step function.*

The results for existence and uniqueness of invariant distributions for a discrete-time Markov process give the exact conditions for existence and uniqueness of the invariant distribution K . We combine the results of Lemmas 1 and 2 and obtain the following proposition.

Proposition 1. *Let $K(t)$ be the invariant (right) cumulative distribution of times between consecutive observations as in equation (4) or in equation (6). The density $q(t)$ of the stationary cross-sectional distribution of times until the next observation is proportional to $K(t)$:*

$$q(t) = RK(t) \text{ for all } t \geq 0 \quad \text{and where} \quad R = \frac{1}{\int_0^\infty K(s) ds}. \quad (7)$$

The right CDF $Q(t)$ is readily found by integration using that $Q(0) = 1$.

For a proof see Appendix A. The results of Proposition 1 substantially generalize the results in the literature about the way the economy aggregates starting from a given individual firm distribution of observation times $H(t|s)$. Existing results in the literature (e.g. Reis, 2006; Carvalho and Schwartzman, 2012) are based on the case of i.i.d. observations, i.e. $H(t|s_1) = H(t|s_2)$ for all s and all pairs (s_1, s_2) . The interpretation of Proposition 1 is simple: in a cross-section there is a mass $K(t)$ of firms who will wait at least t periods before the next observation. The parameter R , measuring the average frequency of observation, scales the density proportionately. Notice that the function $q(t)$ is a density function even in the cases in which the invariant distribution of times between consecutive observations $K(t)$ does not have a density, as in the case where $K(t)$ is a step function (e.g. the Taylor's model discussed above) or, more generally, the case discussed in equation (6).

2.1. Three applications of Proposition 1

Next we discuss three applications of Proposition 1. In the first one, we show that the cross-sectional distribution of times between observations, $K(t)$, is a key determinant of the real effects of monetary shocks. The second application is literature driven: we derive necessary and sufficient conditions under which the firm distribution of observation times $H(t|s)$ yields a distribution $K(t)$ that is both i.i.d. and exponential, a case studied by Reis (2006) and extensively used in applications following Mankiw and Reis (2002). The third application studies the mapping from the distribution $K(t)$ to the size distribution of price changes in a simple model of price dynamics. This mapping can be used to identify the relevant statistics about the distribution $K(t)$ using data on the size distribution of price changes.

2.1.1. The distribution of observations and the effect of monetary shocks. The cross-sectional distribution of the times until the next observation, $Q(t)$, is a key ingredient in the propagation of monetary shocks into prices and output, y_t . We summarize this effect in the single statistic $\mathcal{M}(\delta)$ which measures the real cumulative output effect $\int_0^\infty \log(y_t/y_0) dt$ of a once and for all increase of the aggregate nominal demand of size δ occurring at time $t = 0$:

$$\mathcal{M}(\delta) \equiv \int_0^\infty \log\left(\frac{y_t}{y_0}\right) dt = \frac{\delta}{\epsilon} \int_0^\infty Q(t) dt. \quad (8)$$

The first equality defines the cumulative output effect. The second equality is a result of the following assumptions about the firm's price-setting behaviour and the relationship between prices and output: (1) the price of each firm follows the plan decided before the aggregate shock

until a new observation time. Until the observation occurs, the firm's price is on average δ log points lower than in the absence of the shock. (2) The elasticity of output to nominal demand for those firms that keep prices unchanged is constant and equal to $1/\epsilon$. The statistic $Q(t)$ measures the mass of firms that do not make an observation for at least t periods after the shock, and hence the integral $\int_0^\infty Q(t)dt$ measures the cumulated mass of firms that will not update their prices in the future. This integral ranges between zero, when all firms adjust immediately so that $Q(t)=0$ for all $t \geq 0$, to infinity when firms never adjust so that $\lim_{t \rightarrow \infty} Q(t)=1$. The units are log percentage points, so that a value of $\mathcal{M}(\delta)=0.01$ measures a cumulated increase equal to 1% relative to the steady state output. In the Online Appendix we give more details for this argument and provide references to a general equilibrium setup, the one in Golosov and Lucas (2007), that is consistent with the assumed behaviour. In Table 1, we quantify the cumulated output effect using a calibrated version of the model.

The next proposition shows that $\mathcal{M}(\delta)$ is linear in δ and that it is completely characterized by the mean and variance of the cross-sectional distribution of times between observations:

Proposition 2. *The statistic $\mathcal{M}(\delta) \equiv \frac{\delta}{\epsilon} \int_0^\infty Q(t)dt$ is given by*

$$\mathcal{M}(\delta) = \frac{\delta}{\epsilon} \frac{E_K(t)}{2} (1 + (CV_K(t))^2), \quad (9)$$

where $E_K(t)$ and $CV_K(t)$ are, respectively, the average duration and coefficient of variation of times elapsed between consecutive observations.

For a proof see Appendix A. The linear relation between the size of the monetary shock and the real cumulative output is a distinctive feature of models with time-dependent decision rules which contrasts sharply with the non-linear behaviour that is produced by menu cost models.¹⁰ Both the mean and the variance of the times between consecutive observations have a first-order effect on \mathcal{M} and, therefore, on the real effects of monetary shocks. Notice that the results of Proposition 2 are independent of the assumptions about the primitive distribution of times between observations of the firm, H . A result similar to Proposition 2 has been obtained by Carvalho and Schwartzman (2012) for i.i.d. observation times. In our framework, this corresponds to the case where $H(t|t_1) = H(t|t_2)$ for all t and all pairs (t_1, t_2) . Proposition 2 extends Carvalho and Schwartzman (2012) that allows for observation times to follow a Markov process, a case that arises naturally from the microeconomics of the firm problem (see Sections 4 and 5).

To illustrate the economics of Proposition 2 by means of a simple example consider an economy where each firm draws an observation of length \underline{T} with probability $1-p$, and otherwise an observation of length $\bar{T} > \underline{T}$, independently of the length of the last observation. The distribution of the firm's observation times is given by $\hat{H}(t) \equiv H(t|t')$ for all $t > 0$, where:

$$\hat{H}(t) = \begin{cases} 1 & \text{if } t \in [0, \underline{T}] \\ p & \text{if } t \in (\underline{T}, \bar{T}] \\ 0 & \text{otherwise} \end{cases}$$

Given that observations are i.i.d. over time, the cross-sectional distribution of times between consecutive observations equals the firm's distribution of times between observations. That is,

10. See Alvarez and Lippi (2014) for a theoretical analysis of how the cumulative output effect depends on the size of the monetary shock in menu cost models and Alvarez *et al.* (2015) for an analysis of models with both menu cost and observation cost.

$K(t) = \hat{H}(t)$ for all t (as implied by Lemma 2). Let $\hat{T} \equiv (1-p)\underline{T} + p\bar{T}$ denote the average length of time between consecutive observations. Using Proposition 1 we obtain that $R = 1/\hat{T}$ and the invariant (right) CDF of the cross-sectional times until the next observation is

$$Q(t) = \begin{cases} 1 - t/\hat{T} & \text{if } t \in [0, \underline{T}] \\ 1 - \underline{T}/\hat{T} - p(t - \underline{T})/\hat{T} & \text{if } t \in (\underline{T}, \bar{T}] \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

To understand equation (10) it is useful to compare two example economies. In the first economy, indexed by a subscript 1, $p_1 = 1$ and $\underline{T}_1 = \bar{T}_1 = \hat{T}$. Then $H_1(t)$ is degenerate; the time between consecutive observations is always \hat{T} . This is a case studied, for instance, by Caballero (1989) and Reis (2006). In the second economy, indexed by a subscript 2, $p_2 \in (0, 1)$, so that $\underline{T}_2 < \hat{T} < \bar{T}_2$, and observation times are truly random. As the mean time between consecutive observations is the same in the two economies, the distribution of observation times in economy 2 is obviously a mean-preserving spread of those in economy 1. And so Proposition 2 implies that the more variable times between observations in economy 2 cause larger real effects of a monetary shock.

This simple example clarifies the economics of Proposition 2. Intuitively, $Q_2(t)$ decays more slowly than $Q_1(t)$ (at least for $t > \underline{T}_2$), so that the cumulative effect, as summarized by $\mathcal{M}(\delta)$, is larger. Both functions start from the same point $Q(0) = 1$ but in economy 2 there is a longer tail of possible observation times. This tail boosts the real effect of a monetary shock by delaying the response of some firms, slowing the incorporation of the shock into prices and so prolonging the real effect. It is easy to see this formally either by using equation (9) or integrating $Q(t)$ directly to give $\mathcal{M}(\delta) = \frac{\delta}{\epsilon} (\hat{T}/2) \times [(1/p)(1 - \underline{T}/\hat{T})^2 + (\underline{T}/\hat{T})(2 - \underline{T}/\hat{T})]$. This shows how, for a given average time between observations \hat{T} , the size of the real effects of a monetary shock decreases with p and \underline{T} . Notice that \mathcal{M} is minimized at $p = 1$, where it takes the value of $\hat{T}/2$. This corresponds to economy 1 discussed above. A mean preserving spread of times between consecutive observations increases the real effects of monetary shocks.

2.1.2. The case of exponential i.i.d. observations. The case where observation times are i.i.d. has been first studied by Reis (2006). This case occurs when $H(t|s)$ does *not* depend on s , so we denote it simply by $\hat{H}(t)$. Using Proposition 1 we obtain that the invariant distribution K coincides with H and thus:

$$q(t) = R\hat{H}(t) \text{ for all } t \geq 0, \quad (11)$$

where the number of observations per unit of time is given by $R = 1/\int_0^\infty \hat{H}(t)dt$.¹¹ An immediate implication of equation (11) is that:

Corollary 1. *Let the observation times be i.i.d. so that $H(t|s) = \hat{H}(t)$ for each $s \geq 0$. The cross-sectional distribution of the times until the next observation, Q , is exponential if and only if the firm's distribution of the times between observations, \hat{H} is exponential.*

The proof follows since an exponential density is proportional to its right CDF, a property shared by $q(t) \propto \hat{H}(t)$ in equation (11). Corollary 1 clarifies an important result in Reis's (2006),

11. Note that this is the case considered for a renewal process, where Q is the forward recurrent time as, for instance, in Karlin and Taylor (1998).

namely his Proposition 6, in which the result is obtained under the same assumptions of this section about the distribution H (i.e. firms observation times being i.i.d. so that $H(t'|t) = \hat{H}(t)$ for all t' and t), but where the requirement that the distribution H has to be itself exponential in order for the cross-sectional distribution Q to be exponential is not as explicit.¹² This clarification is important because the case of an exponential cross-sectional distribution Q has been extensively used in the “rational inattentiveness” literature that followed the seminal work by Mankiw and Reis (2002). We conclude that the conditions for an exponential cross-sectional distribution Q are rather restrictive: $\hat{H}(t)$ can only be exponential. Section 5.1 below will further explore this case by establishing what primitive assumptions about the firm information and the distribution of observation costs are necessary for the firm distribution of observation times to be exponential.

2.1.3. From the distribution of observations to the distribution of price changes.

Given the result of Proposition 2 we are particularly interested in identifying the variability of observation times. In this section, we consider a special case where the cross-sectional distribution of times between consecutive observations, K , can be identified from observable statistics on the distribution of the size of price changes. We derive this mapping assuming that the target price, p_t^* , evolves according to a Brownian motion, $d\log(p_t^*) = \mu_p dt + \sigma dB_t$, where B_t is the realization of an idiosyncratic Wiener process, and $\mu_p = -\sigma^2/2$ implies that p_t^* is a martingale.¹³ In this specification of the model the cross-sectional invariant distribution of log-price changes, $\Delta\log(p)$, is given by a mixture of normals indexed by t , where the mixture has density corresponding to the cross-sectional distribution of times between consecutive observations, i.e. $k(t) = -K'(t)$, and each of the normals has mean and variance $(0, \sigma^2 t)$. The next lemma characterizes two useful moments of the size distribution of price changes (see Appendix A for the proof):

Proposition 3. *Let the target price p_t^* be a martingale evolving according to a Brownian motion, $d\log(p_t^*) = -\sigma^2/2 dt + \sigma dB_t$, then the variance of the log-price changes is*

$$\text{Var}(\Delta\log(p)) = \sigma^2 E_K(t) \quad (12)$$

where $E_K(t) = \int_0^\infty K(s) ds = 1/R$ is the average duration of time between observations, and the kurtosis of the log-price changes is equal to:

$$\text{Kurt}(\Delta\log(p)) = 3 \left[(CV_K(t))^2 + 1 \right], \quad (13)$$

where $CV_K(t)$ is the coefficient of variation of times between consecutive observations (taken with respect to the distribution K).

Proposition 3 is useful for two reasons. First, combined with the results of Lemma 2, it allows us to obtain statistics about the cross-section distribution of price changes from the primitive assumptions about the individual firm distribution of times between observations H . Secondly, combined with observable statistics on the distribution of price changes, it allows us to identify the coefficient of variation in observation times from the kurtosis of the distribution of price changes,

12. In private correspondence Reis clarified to us that his result follows from his definition of stationarity in Reis’s (2006) Definition 1.

13. In Section 3, we discuss the possibility of price plans and give the conditions under which p_t^* is a martingale. We argue that this case is not knife edge and that its analysis is informative for environments in which p_t^* has a small drift (as is the case with a small inflation or with a productivity trend).

as well as to identify the volatility of the target price from the variance and frequency of price changes. For instance, if we use a kurtosis of price changes equal to 4, consistent with the estimates by Alvarez *et al.* (2014), we obtain a coefficient of variation equal to $CV_K(t) = \sqrt{1/3}$. Combining such estimate for $CV_K(t)$ with the results of Proposition 2, we obtain that the cumulative real effects of a permanent innovation to money supply is about 35% larger than the corresponding figure predicted by models with deterministic observation times (where $CV_K(t) = 0$, such as Caballero, 1989 and Reis, 2006), and about one-third smaller than predicted by models with exponential observation times (where $CV_K(t) = 1$, such as Mankiw and Reis, 2002).

The results of Proposition 3 (but not the other results of this section) depend on the assumption that the innovations in $\log(p_t^*)$ are Gaussian. In the Online Appendix, we extend our analysis to a model where the innovations in $\log(p_t^*)$ are not Gaussian because of the random volatility of those innovations. This extension is useful to study the mapping from the distribution of price changes to the distribution of observation times in a framework where the variability of observation times is driven by a process with stochastic volatility.

3. A FIRM PROBLEM LEADING TO OPTIMAL OBSERVATION TIMES

This section studies a model of the firm's optimal decision for observation times. This problem gives rise to a Markov process for observations that can be represented by the conditional distribution $H(t'|t)$, the primitive of our analysis in Section 2. We set up the firm's decision problem in a relatively general formulation that encompasses several cases of interest: one where the duration of the optimally chosen observation times is constant, one where it follows a first-order Markov process (Section 4), and one where it is i.i.d. (Section 5). The distinction between models of constant observation times (*e.g.* Caballero) and models of random observation times is important since the variability of observation times increases the real effects of monetary shocks, as shown in Proposition 2. Within the class of models that deliver variation in observation times, the case of Markov persistent durations of the time between observations is appealing both because we think it is a realistic representation of firms' behaviour and because, everything else being equal, more persistent durations of the time between observations increase the real effects of monetary shocks.

While the aggregation results of Proposition 1 apply to any model of the firm's problem that can be represented by a distribution $H(t'|t)$, we focus on a specific framework where variation in observation times arises from variation in observation costs. This framework allows us to study two cases of interest in the literature. First, the existing literature has provided a characterization of Markov observation times, but only under the assumption of negligible observation costs (*e.g.* Reis, 2006). We provide a global characterization of the decision rule, which allows for both negligible as well as non-negligible observation costs. The latter are essential to get infrequent observations. Secondly, a large literature has focused on i.i.d. observation times and its implications for macroeconomic dynamics as discussed above. Despite its widespread use, the "rational inattentiveness" literature has not provided a characterization of optimal i.i.d. observations originating from random variation in the benefit/cost of acquiring information. This article fills that gap and shows what features of the primal problem are relevant to obtain i.i.d. distributed observation times. This section sets up a general environment where the firm's optimal stopping times can be studied. Section 4 discusses the general case of Markovian observation times. Section 5 specializes the model to obtain i.i.d. observation times, the standard in the literature.

The underlying economy is a simple variant of Golosov and Lucas (2007). Firms make two types of decisions: (1) when to acquire information, and (2) the price at which to sell their good. In particular, there is a continuum of firms, each producing a variety denoted by i , and facing

a downward sloping demand given by $C(P_{i,t}/W_t) = A (P_{i,t}/W_t)^{-\eta}$, where $A > 0$ is a constant, $\eta > 1$ is the price-elasticity, $P_{i,t}$ is the nominal price of variety i , and W_t is the nominal wage in period t . The production technology of each variety is linear in labour with idiosyncratic labour productivity $z_{i,t}$. It is well known that in this environment the frictionless optimal pricing strategy of each firm is given by a constant markup over nominal marginal cost, given by $W_t/z_{i,t}$. The (log) of the labour productivity follows a Brownian motion with drift ϕ and volatility σ : $d\log(z_{i,t}) = \phi dt + \sigma dB_{i,t}$, where $B_{i,t}$ is the realization of an idiosyncratic Wiener process. We also assume that the product of the firm is replaced (“dies”) with a constant probability per unit of time equal to λ , and a new product is born with productivity $z = 1$. From a modelling point of view the exogenous substitutions allow for an invariant distribution of firm’s productivities and for bounded output. Finally, we assume that the nominal wage grows at a constant rate μ , so that its dynamics between t and $t + T$ are given by $W_{t+T} = W_t e^{\mu T}$.

Information gathering: In the spirit of the rational inattentiveness literature we assume that to gather information about the nominal marginal cost the firm must pay a fixed information-gathering cost (see *e.g.* Reis, 2006), an activity we refer to as an *observation*. Hence, at observation times, there is a discrete change in the information about the nominal marginal cost. While acquiring the information is costly, firms can change prices at any time at no cost.¹⁴

The value of information: We first develop notation to quantify the value of information. For expositional simplicity, we drop the time and firm sub-indexes. The firm’s period profits, scaled by the nominal wage, are given by

$$\Pi(p, z) \equiv C(p) \left(p - \frac{1}{z} \right), \quad \text{where } p \equiv \frac{P}{W}, \quad (14)$$

so that real profits are a function of the ratio of the nominal price to the nominal wage, p , and of productivity, z .¹⁵

When prices are set conditional on perfect information of the current productivity z the expected profits are

$$\Pi^*(z) \equiv \max_p \Pi(p, z) = A \left(\frac{1}{z(\eta - 1)} \right)^{1-\eta} \eta^{-\eta}, \quad \text{with maximizer } p^*(z) = \frac{\eta}{\eta - 1} \frac{1}{z}$$

where $p^*(z)$ denotes the instantaneous profit-maximizing price as a function of productivity. We notice that $p^*(z)$ is the standard markup pricing of monopolistic competition.

To highlight the role of imperfect information, we compare the T -periods ahead expected profits under perfect information with the expected profits under imperfect information. Let $L(\cdot; T|z)$ denote the CDF of T -periods ahead productivity conditional on the current productivity, z . Given the assumptions on the process for z , $L(\cdot; T|z)$ is a log-normal distribution. In the case of perfect information, the T -periods ahead expected profit evaluated conditional on current productivity z and the optimal pricing function $p^*(z')$, is given by

$$\int_0^\infty \Pi^*(z') dL(z'; T | z) = \Pi^*(z) e^{bT}, \quad \text{where } b \equiv (\eta - 1)(\phi - \sigma^2/2) + \eta(\eta - 1)\sigma^2/2. \quad (15)$$

14. This differs from the models in which firms face a joint cost of observing the state and adjusting the price as in Bonomo and Carvalho (2004), where the two costs are associated, or Alvarez *et al.* (2011) where the costs are dissociated and the firm can select which one to incur.

15. We notice that, given the deterministic path of the nominal wage, scaling profits by the nominal wage and studying the firm problem with respect to the scaled price p is without loss of generality with respect to the information set of the firm.

Alternatively, in the case of imperfect information the expected profits T periods ahead, conditional on current productivity z , is a function of the best estimate of productivity z' :

$$\hat{\Pi}(T, z) \equiv \max_p \int_0^\infty \Pi(p, z') dL(z'; T | z), \quad (16)$$

with maximizer $\hat{p}(T, z)$ given by

$$\hat{p}(T, z) = \frac{\eta}{\eta - 1} \int_0^\infty \frac{1}{z'} dL(z'; T | z) = \frac{\eta}{\eta - 1} \frac{1}{z} e^{(\sigma^2/2 - \phi)T}. \quad (17)$$

The optimal price $\hat{p}(T, z)$ is thus given by the markup $\eta/(\eta - 1)$ times the T periods ahead expected value of the marginal cost $1/z'$ (normalized by wages), conditional on the last observed productivity z . We notice that $\hat{p}(0, z) = p^*(z)$. Using equation (16) into equation (14) we can write $\hat{\Pi}(T, z)$ as

$$\hat{\Pi}(T, z) = \Pi^*(z) e^{aT}, \quad \text{where } a \equiv (\eta - 1)(\phi - \sigma^2/2). \quad (18)$$

The two expected profits in equation (15) and equation (18) are taken with respect to the same T -periods horizon. The difference between the two expressions lies in the information used to set prices. When prices are set based upon T -periods-old information the expected profits grow at the rate a . Instead, when prices are set with complete information the expected profits grow at the rate $b > a$. Hence $B \equiv b - a = \eta(\eta - 1)\sigma^2/2$ captures the rate at which the benefit of acquiring information increases as a function of the time elapsed since the last observation.¹⁶ Notice that the value of information equals (one-half) the curvature in the profit function, $\eta(\eta - 1)$, times the incremental uncertainty σ^2 . Intuitively, a higher the volatility of productivity innovations (σ) lowers the information content of past observations, increasing the value of information. Likewise, a higher demand elasticity (η) boosts the impact of a given error in pricing on the firm's profits, thus increasing the value of information.

Price plans and price changes: Price changes between observation dates are referred to in the literature as *price plans*: such price changes are based upon the information gathered in the last observation and the law of motion of the relevant states. Equation (17) illustrates the workings of price plans in our model: at each time T following the last observation, the firm charges the price that maximizes expected profits. The nominal price of the firm grows at the rate $\mu + \sigma^2/2 - \phi$, where μ is the component reflecting nominal wage growth (recall that we are using $p = P/W$ as the control variable), and the $\sigma^2/2 - \phi$ reflects the expected growth of productivity. Thus, if $\mu - \phi + \sigma^2/2 \neq 0$ then the optimal price plan implies that prices change continuously. This is a common element in models of rational inattentiveness that lack a physical cost of price adjustment. A robust pattern in the data is, however, that prices change infrequently. A simple way to obtain infrequent price changes in this class of models is to assume that the level of the nominal marginal cost is a martingale, *i.e.* $\mu - \phi + \sigma^2/2 = 0$. As a result, price changes only occur when new information arrives, so that the frequency of price changes coincide with the frequency of observations. While this distinction is immaterial for the theoretical results of this article, being able to map this class of models to statistics about the size and frequency of price changes may be relevant for their quantitative implications, as we already showed in Section 2.1.3. Moreover,

16. Abel *et al.* (2007) also solve for optimal observation times in a framework where the benefit of doing an observation increases at a constant rate as a function of time elapsed.

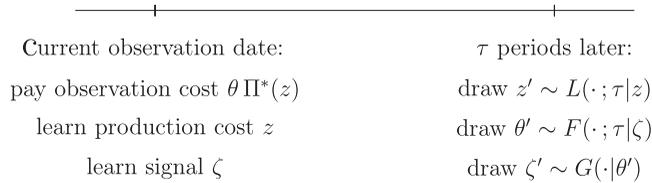


FIGURE 1
Time line

in Alvarez *et al.* (2011) we show that price plans would not be optimal even in the presence of a drift in the nominal marginal cost, when a price adjustment cost is added to a similar model and calibrated to match the frequency of price changes in the U.S. economy. Therefore, in our quantitative applications we will focus on a calibration of the model where the nominal marginal cost is a martingale.

The cost of gathering information: We assume that acquiring information about the level of the production cost $1/z$ requires the payment of an observation cost. We interpret the observation cost as the physical cost of acquiring the information needed to make the price decision as well as the time cost of decision making in the firm (gathering and aggregating information, *e.g.* Zbaracki *et al.*, 2004; Reis, 2006). The firm incurs an observation cost any time it acquires information about the state. Upon payment of the observation cost, the firm learns the current value of z . No information on the state arrives between observation dates.¹⁷ We assume that the observation cost is equal to $\theta \Pi^*(z)$, so that it is proportional to the value of the static monopolist profit under complete information. This assumption serves two purposes. First, it reduces the technical difficulty of the problem: the two expected profits in equations (15) and (18) show that the benefits of information are proportional to $\Pi^*(z)$. Making the costs scale similarly leaves the entire firm’s problem homogeneous in $\Pi^*(z)$, which simplifies the analysis. Secondly, when z is persistent, this assumption is also necessary to produce i.i.d. times between observations as we will do in Section 5.

We will call θ the observation cost. We allow for θ to be random, with a finite lower bound $\underline{\theta}$. The processes of the observation cost θ and of the production cost z are assumed to be independent. Immediately after paying the observation cost, the firm not only learns the current value of z , but also receives a signal $\zeta \in [\underline{\zeta}, \infty)$ which is informative about the future realizations of θ . In particular, the signal ζ summarizes all the information about the value of the observation cost to be paid τ periods from now, for any $\tau \geq 0$. Mathematically we write $F(\theta'; \tau|\zeta)$ to be the CDF of the observation cost $\theta' \in [\underline{\theta}, \infty)$ to be paid τ periods after the current observation, conditional on the signal ζ . The dependence of the distribution F on τ allows the distribution of the observation cost θ' to vary with the time elapsed between observations. In practice, we will work with a specification where the signal ζ is proportional to the expected cost of the next observation for any given τ , *i.e.* $E[\theta'; \tau|\zeta] \propto \zeta$. Upon the next observation, when a particular cost θ' is realized, a new signal $\zeta' \in [\underline{\zeta}, \infty)$ is drawn from the CDF $G(\cdot|\theta')$. The timeline in Figure 1 describes the structure of the observation cost θ , the associated signal ζ and production cost z , for an observation date that occurs τ periods after the current observation.

17. Notice the difference with the “rational inattention” literature that developed after Sims (2003) where agents typically process a limited amount of information every period. See Mackowiak and Wiederholt (2009) for an application to firm price setting.

The functions F and G fully characterize the process for the observation cost, and provide enough flexibility to cover cases discussed in the literature as well as generalizations that we find useful. For instance, the model of deterministic observation times studied by Caballero (1989) and Reis (2006) is encompassed by our framework if the signal is uninformative about the future observation cost, which is the case if $F(\theta', \tau_0 | \zeta_0) = F(\theta', \tau_1 | \zeta_1)$ for all θ' and all pairs (τ_0, ζ_0) . In this case, the distribution G is irrelevant because, given that the signal is uninformative, the mechanism to obtain the new signal is irrelevant.¹⁸

Another case discussed in the literature is one where the firm's observation times are i.i.d., as proposed by Reis (2006). Our model provides a foundation to i.i.d. observation times: the firm has to draw a signal about the future observation cost that is both informative about the next observation cost and independent of all other shocks (including the current value of the observation cost). In this case, the particular form of the distribution G is relevant. Formally, observation times are i.i.d. in our model if and only if $G(\zeta | \theta_0) = G(\zeta | \theta_1)$ for all ζ and all pairs θ_1, θ_0 . The distribution F shapes the precision of the signal. Finally, the more general case where $G(\zeta | \theta_0) \neq G(\zeta | \theta_1)$ for at least some ζ and some pairs $\theta_1 \neq \theta_0$ allows us to extend our analysis to the case of observation times correlated over time, a case which we find more reasonable than the i.i.d. assumption.

The firm's Bellman equation: Without loss of generality we consider the problem of the firm right after it has paid the observation cost, when it is deciding the length of the time until the next observation which we denote as τ . The state of the firm at this date consists of both the signal ζ and the productivity z . The value of the firm $V(\zeta, z)$, *i.e.* the expected discounted value of the profits net of the observation cost, solves the following recursive problem

$$V(\zeta, z) = \max_{\tau \in \mathbb{R}_+} \int_0^\tau e^{-(\rho+\lambda)s} \hat{\Pi}(s, z) ds \quad (19)$$

$$+ e^{-(\rho+\lambda)\tau} \int_{-\infty}^{\infty} \int_{\underline{\theta}}^{\infty} \left(-\theta' \Pi^*(z') + \int_{\underline{\zeta}}^{\infty} V(\zeta', z') dG(\zeta' | \theta') \right) dF(\theta'; \tau | \zeta) dL(z'; \tau | z),$$

where real profits are discounted at rate $\rho + \lambda$, with ρ capturing the agent's preferences discount rate in the setup of interest (described in the appendix of Alvarez and Lippi, 2014), and λ is due to the substitution of products. The firm chooses the time $\tau \geq 0$ between the current and the next observation, where $\tau = \infty$ represents the case of never observing. The first term on the right-hand side of equation (19) reflects the expectations of cumulated profits until the next observation. Equations (15)–(18) imply that the s -periods ahead expected profits, $\hat{\Pi}(s, z)$, scale with $\Pi^*(z)$, *i.e.* $E[\theta_\tau \Pi^*(z_\tau) | \zeta_0, z_0] = E[\theta_\tau | \zeta_0] e^{b\tau} \Pi^*(z_0)$. The second term on the right-hand side of equation (19) refers to the discounted expectation of the continuation value upon the next observation, $V(\zeta', z')$, conditional on the next observation date taking place τ periods from now, the current productivity being z , and the current signal being ζ . The integrals on the right-hand side of equation (19) reflect the expectation of $V(\zeta', z')$ with respect to the possible realizations of z' and ζ' (which also depend on the realization of θ').

We use that both expected profits and observation cost scale with $\Pi^*(z)$, together with the independence of z' and θ' , to express the value function as $V(\zeta, z) = v(\zeta) \Pi^*(z)$, where the function

18. In Caballero (1989) and Reis (2006) the cost is fixed, which amounts to a degenerate F . Our setup thus delivers a small generalization: absent an informative signal, observation times will be constant even if F is not degenerate.

$v(\zeta)$ solves the following simpler recursive problem:

$$v(\zeta) = \max_{\tau \in \mathbb{R}_+} \int_0^\tau e^{-(\rho+\lambda-a)t} dt + e^{-(\rho+\lambda-b)\tau} \int_{\underline{\theta}}^\infty \left(\int_{\underline{\zeta}}^\infty v(\zeta') dG(\zeta'|\theta') - \theta' \right) dF(\theta'; \tau|\zeta). \tag{20}$$

We denote by the function $\tau(\zeta)$ the optimal policy that solves the firm’s problem. Notice that this problem has relatively few parameters. The functions $v(\cdot)$ and $\tau(\cdot)$ depend only on three scalars (a , b and $\rho + \lambda$), and two functions (G and F). The first term on the right-hand side of equation (20) contains the value of expected profits using only current information. The second term contains the value of expected profits after a new observation, as well as the expected future observation cost. The optimal choice of τ balances the value of information (summarized by $b > a$) against the expected cost of a new observation θ' . An immediate result of equation (20) is that the optimal time between observations only depends on the realization of the signal ζ , and not on productivity z . Finally, to have a finite value of a firm in the frictionless case, we assume that $\lambda + \rho > b$. Hence, the value function is bounded below and above as follows:

$$0 < \underline{v} \equiv \frac{1}{\rho + \lambda - a} \leq v(\zeta) \leq \bar{v} \equiv \frac{1}{\rho + \lambda - b} < \infty \text{ for all } \theta \in (0, +\infty),$$

where the lower bound is obtained by setting $\tau = \infty$, while the upper bound is obtained in the perfect information case when observation is always costless, and the firm observes continuously.

Obtaining $H(t'|t)$: We can use the solution to the firm’s problem $\tau(\zeta)$, and the properties of the process governing the signal ζ , to derive the conditional right CDF for the times between consecutive observations, which in Section 2 was denoted by $H(\cdot|t)$. In particular, assume that $\tau(\cdot)$ is strictly increasing in ζ , a property which will be satisfied in our applications, and let $\zeta = \hat{\zeta}(t)$ denote the mapping from an observation of length t with the associated signal ζ , where $\hat{\zeta}(t) = \tau^{-1}(t)$ is the inverse function associated with the optimal policy $\tau(\cdot)$. For simplicity, consider the case with no substitutions, *i.e.* $\lambda = 0$.¹⁹ In this case, the distribution H is given by $H(t'|t) = \int_{\underline{\theta}}^\infty (1 - G(\hat{\zeta}(t')|\theta')) dF(\theta'; t|\hat{\zeta}(t))$, which equals the probability of drawing a signal $\zeta' \geq \hat{\zeta}(t')$ at the next observation, conditional on having drawn a signal $\zeta = \hat{\zeta}(t)$ at the last observation, which occurred t periods ago. This requires integrating over all possible realizations of the future observation cost θ' , since the realization of θ' determines the distribution from which the signal ζ' will be drawn upon the next observation.

3.1. The case of constant optimal observation times

This subsection presents a specification of the observation cost process which implies that the optimal (consecutive) observation times have equal duration. This case serves as a useful benchmark, and is a limiting case, of the more general case of random observations that will be considered in the next sections. In particular, we assume that signals are uninformative about future observation costs, so that firms face the same problem upon every observation and choose the same time to the next observation. We assume that $F(\theta, \tau_0|\zeta_0) = F(\theta, \tau_1|\zeta_1) \equiv \hat{F}(\theta)$ for all θ and all pairs (τ_0, ζ_0) and (τ_1, ζ_1) and that G is arbitrary. The assumption on F means that the signal is uninformative about future observation costs. Note that the signal could be uninformative and

19. The case with substitutions follows a similar logic, but is algebraically more involved (see the Online Appendix).

yet $\hat{F}(\theta)$ does not need to be degenerate. The function G can be left unspecified because, given that the signal is uninformative, the mechanism to obtain the new signal is irrelevant. Using these assumptions on F in equation (20), one obtains that $\tau(\zeta) = \hat{\tau}$ and $v(\zeta) = \hat{v}$, so neither function depends on ζ . In particular, letting the expected observation cost $E[\theta] = \int_0^\infty \theta d\hat{F}(\theta) > 0$, the Bellman equation becomes:

$$\hat{v} = \max_{\tau \in \mathbb{R}_+} \int_0^\tau e^{-(\rho+\lambda-a)t} dt + e^{-(\rho+\lambda-b)\tau} (\hat{v} - E[\theta]) . \quad (21)$$

The analysis that follows characterizes the comparative statics of the optimal decision rule, summarized by the constant $\hat{\tau}$, as a function of the model parameter $E[\theta]$, *i.e.* the expected value of the observation cost. To do so, as well as for future use, we define the function $\tilde{\tau}(\cdot)$ so that the optimal policy is $\hat{\tau} = \tilde{\tau}(E[\theta])$. The function $\tilde{\tau}(\cdot)$ is

$$\tilde{\tau}(x) = \begin{cases} \frac{1}{B} [\ln(\bar{v}) - \ln(\bar{v}(x) - x)] & \text{if } 0 < x < \underline{v} \\ \infty & \text{if } x \geq \underline{v} \end{cases} , \quad (22)$$

$$\text{where } \bar{v}(x) = \begin{cases} \text{smallest solution of } \bar{v} = \underline{v} + B\underline{v}\bar{v} \left(\frac{\bar{v}-x}{\underline{v}} \right)^{\frac{1}{B\underline{v}}} & \text{if } 0 < x < \underline{v} \\ \underline{v} & \text{if } x \geq \underline{v} \end{cases} , \quad (23)$$

and $B \equiv b - a = \eta(\eta - 1)\sigma^2/2$ summarizes the value of information. The following proposition collects these results (see Appendix A for the proof):

Proposition 4. *Assume that signals are not informative, *i.e.* $F(\theta, \tau_0 | \zeta_0) = F(\theta, \tau_1 | \zeta_1) \equiv \hat{F}(\theta)$ for all θ and all pairs (τ_0, ζ_0) and (τ_1, ζ_1) , and $E[\theta] > 0$. The optimal observation time is given by $\hat{\tau} = \tilde{\tau}(E[\theta])$, which is strictly increasing in $E[\theta]$ for $0 < E[\theta] < \underline{v}$, and infinite for $E[\theta] \geq \underline{v}$.*

The optimal time between observations is increasing in the expected observation cost: the larger the expected observation cost, the more time must elapse before the benefit of an observation is large enough to induce the firm to pay the observation cost. As the expected value of the observation costs becomes small, formally as $E[\theta] \downarrow 0$, the optimal observation time is approximated by the square root function $\hat{\tau} \approx \sqrt{2E[\theta]/B}$. This limiting case in which $E[\theta] \downarrow 0$ reproduces Reis's (2006) in Proposition 5, where a perturbation method was used to derive an approximation for $\hat{\tau}$. Proposition 4 generalizes Reis's results by characterizing the solution away from the case where $E[\theta]$ is small. These results will be useful in the next section to characterize the determinants of the shape of the optimal observation times in general as a function of the expected value of the observation cost shocks and of their persistence.

4. OPTIMAL MARKOV OBSERVATION TIMES

The results of Proposition 2 imply that the case of constant observation times studied in the literature and discussed in the previous section predicts the smallest real effects of monetary shocks for given average time between observations $E_K(\tau) = \hat{\tau}$, as the associated coefficient of variation of observations is $CV_K(\tau) = 0$. Studying the underpinnings of variability in observation times is important because such variability increases the size of the real effects of monetary shocks. This section presents a specification of the process for the observation cost implying that optimal observation times form a first-order Markov process. The solution to the firm's problem

will provide a mapping from the realization of the observation cost to the optimal time between consecutive observations, τ .

We will use this mapping to obtain a relationship from the coefficient of variation of observation costs, $CV(\theta)$, to the coefficient of variation of observation times, $CV_K(\tau)$. This mapping is useful to appreciate the ability of the model to generate a given variation in observation times, for a given variation of the observation costs.

4.1. Solving for the observation time with a specific cost distribution

We assume that the observation cost $\theta \geq 0$ follows a continuous time process, and the firm learns about the current value of θ only at observation times. The observation cost is modelled in a way that parallels the production cost. Both of them follow a continuous-time stochastic process whose distribution is independent of whether and when the firm decision to observe.²⁰ A special case of the process for the observation cost is the one where the observation cost is constant until a new value is drawn from the distribution $\hat{F}(\cdot)$ with support on $(0, \infty)$. The cumulative distribution in this case is given by

$$F(\theta', \tau | \theta) = e^{-\frac{\tau}{\kappa}} \mathbf{1}_{\theta' \geq \theta} + (1 - e^{-\frac{\tau}{\kappa}}) \hat{F}(\theta') \quad , \quad \kappa < \infty . \quad (24)$$

The arrival rate of the new value is given by the constant $1/\kappa$ (*i.e.* it is a Poisson process), and it is independent of the current value of θ . We will refer to κ as the parameter governing the persistence of observation cost, with a higher κ associated with a higher persistence. We assume that the current realization of the observation cost is the best and only predictor of future realizations of the observation cost. Formally we assume that $\zeta = \theta$, implying $G(\zeta | \theta) = \mathbf{1}_{\zeta \geq \theta}$. Without loss of generality we can replace ζ with θ as the argument of the firm policy and value function, *i.e.* $\tau(\theta)$ and $v(\theta)$.

Given the process for θ implied by equation (24), we can rewrite the Bellman equation in equation (20) as:

$$v(\theta) = \max_{\tau \in \mathbb{R}_+} \int_0^\tau e^{-(\rho+\lambda-a)t} dt + e^{-(\rho+\lambda-b)\tau} \left(e^{-\frac{\tau}{\kappa}} (v(\theta) - \zeta) + \left(1 - e^{-\frac{\tau}{\kappa}}\right) (E[v] - E[\theta]) \right) , \quad (25)$$

where $E[v] \equiv \int_0^\infty v(\theta) d\hat{F}(\theta)$ and $E[\theta] \equiv \int_0^\infty \theta d\hat{F}(\theta)$. The next two propositions characterize the shape of the decision rule $\tau(\theta)$ that solves the problem in equation (25): the general message is that the shape crucially depends on two parameters, the expected value, $E(\theta)$, and the persistence, κ , of the observation cost.

As shown by Reis's (2006) proposition 4 the decision rule, $\tau(\theta)$, is a square root formula for any given level of persistence κ as the expected value of the observation costs becomes tiny, *i.e.* as $E(\theta) \rightarrow 0$. The next proposition extends that characterization (see Appendix A for the proof). It is shown that the shape of the optimal decision may be very different from the square root provided that the persistence of the shocks is sufficiently small. In particular, for a given expected value of the costs, $E(\theta) > 0$ (arbitrarily small), there is a threshold for the persistence $\underline{\kappa}$ such that if $\kappa < \underline{\kappa}$ the firm chooses a strictly positive time until the next observation even if the realized observation cost θ is tiny. Conversely, the square root result is obtained if the observation cost is sufficiently

20. The assumption that the evolution of the observation costs is independent of the decision to observe will need to be abandoned to generate i.i.d. observation times, as discussed in Section 5.

persistent, *i.e.* $\kappa > \bar{\kappa}$. The proposition assumes a sufficiently small level of the expected future costs, *i.e.* for $E[\theta] < \underline{v}$ (if the condition is not satisfied, the optimal time between observations diverges):

Proposition 5. *Assume that $E[\theta] < \underline{v}$, and let $\bar{\kappa} \equiv \underline{v} + E[\theta]/(\bar{v}B)$ and $\underline{\kappa} \equiv E[\theta]/(B\underline{v}) < \bar{\kappa}$:*

- (i) *for any $0 < \kappa < \infty$ we have $\tau(\theta) > 0$ and $\tau'(\theta) > 0$ if $\theta > 0$, with $\lim_{\theta \rightarrow \infty} \tau(\theta) = \infty$ and $\lim_{\theta \rightarrow \infty} \tau'(\theta) = 0$;*
- (ii) *for all $\kappa < \underline{\kappa}$ we have $\lim_{\theta \rightarrow 0^+} \tau(\theta) > 0$;*
- (iii) *for all $\kappa \geq \bar{\kappa}$ we have $\lim_{\theta \rightarrow 0^+} \tau(\theta) = 0$ and $\lim_{\theta \rightarrow 0^+} \tau'(\theta) = +\infty$.*

The first point of the proposition states that the time between consecutive observations is strictly positive provided that $\theta > 0$, since the firm would face unbounded losses in a finite period of time if it were to observe the state continuously. Moreover, point (i) also states that the optimal observation time is strictly increasing in the current value of the observation cost θ , since a higher θ signals a higher expected value of the future observation cost.²¹

The next two points of the proposition are crucial to see how the shape of the optimal rule at small values of θ depends on the persistence of the cost shocks, κ . Point (ii) states a sufficient condition on the persistence parameter κ such that if the persistence is low enough (*i.e.* $\kappa < \underline{\kappa}$) then the optimal time between observations converges to a strictly positive value as the observation cost converges to zero. Intuitively, as the persistence of observation costs decreases, the current observation cost becomes a bad predictor of future observation costs. When the persistence is sufficiently low, such expectation will be dominated by the unconditional mean, *i.e.* $E[\theta] > 0$, to the point that even when the observation cost is tiny the firm will still choose to wait for a strictly positive time before making another observation.

A different rule obtains if the persistence parameter κ is sufficiently high (*i.e.* $\kappa \geq \bar{\kappa}$), a condition described under point (iii). In this case, the optimal time between observations converges to zero as the observation cost converges to zero (with an infinite slope at zero). The reason is that when persistence is high the current observation cost is a good predictor of the future cost, thus the decision rule becomes responsive to the current cost.

Figure 2 illustrates the results of our proposition through numerical examples obtained at parameter values that reproduce the average frequency and size of price changes estimated by Nakamura and Steinsson (2008). In these examples, and in the ones that will follow later in the article, we will concentrate on the case in which the nominal marginal cost is a martingale (*i.e.* $\mu - \phi + \sigma^2/2 = 0$) because, as it is well documented, prices change infrequently.²² We provide more details on the calibration of the model in the Online Appendix.

4.2. Two limiting cases

To further illustrate the key role of the persistence of observation costs for optimal observation times, the next proposition considers two limiting cases for κ : permanent observation cost, *i.e.* $\kappa \rightarrow \infty$; and i.i.d. observation cost, *i.e.* $\kappa \rightarrow 0$ (see Appendix A for the proof). This proposition is informative about the shape of the decision rule, not just at small values of θ as done before,

21. This can be seen noting that $E[\theta'; \tau|\theta] = e^{-\frac{\tau}{\kappa}}\theta + (1 - e^{-\frac{\tau}{\kappa}})E[\theta]$ increases with θ for any time τ .

22. This is a meaningful benchmark also because if one were to add a separate cost of adjusting prices, the optimal policy would indeed feature constant prices between consecutive observations as long as the drift in the marginal cost is not too large (see Alvarez, 2011 for a rigorous characterization).

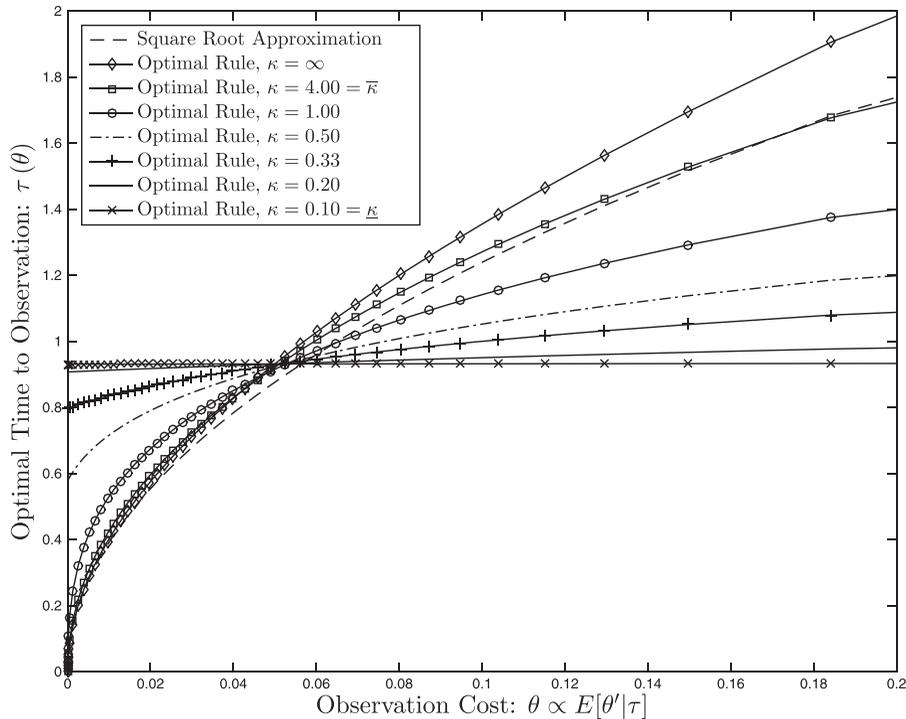


FIGURE 2

The optimal observation times rule $\tau(\theta)$, as persistence κ varies

Note: The figure reports the optimal time to the next observation $\tau(\theta)$, in years, as a function of the signal value $\zeta = \theta$ on the horizontal axis, and for different values of persistence, as measured by κ . The distribution \hat{F} is assumed to be exponential with average observation cost $E[\theta] = 0.05$. Depending on persistence κ , this cost implies between 1.1 and 1.4 observations per year on average. We set the drift $\mu - \phi + \sigma^2/2 = 0$ to obtain infrequent price changes, and $\sigma = 0.114$ to match an average size of price changes equal to 10%. The other parameters are $\eta = 5$, $\lambda = 0.25$, and $\rho = 0.02$. For more details see the Online Appendix. The dashed line plots the function $\sqrt{2\theta/B}$.

showing that it converges pointwise to the decision rule in the problem with constant time between observations when $\kappa \rightarrow \infty$.

Proposition 6. Assume that $E[\theta] < \underline{v}$, and let $\tau(\theta; \kappa)$ and $v(\theta; \kappa)$ denote, respectively, the solution to the firm problem and associated value function in equation (25) conditional on a given persistence parameter κ . Then we have that:

- (i) for any $\theta \geq 0$, $\lim_{\kappa \rightarrow \infty} \tau(\theta; \kappa) = \tilde{\tau}(\theta)$ and $\lim_{\kappa \rightarrow \infty} v(\theta; \kappa) = \tilde{v}(\theta)$ where the functions $\tilde{\tau}(\cdot)$ and $\tilde{v}(\cdot)$ are given by equations (23)–(22);
- (ii) for any $\theta > 0$, $\lim_{\kappa \rightarrow 0} \frac{\partial \tau(\theta; \kappa)}{\partial \theta} = \frac{\partial v(\theta; \kappa)}{\partial \theta} = 0$.

Point (i) shows that when the observation cost is permanent, *i.e.* $\kappa \rightarrow \infty$, the optimal observation time is given by the function $\tilde{\tau}(\theta)$, in equation (22). One can use this expression to show that the optimal policy is approximated by $\tilde{\tau}(\theta) = \sqrt{2\theta/B} + o(\theta)$, implying an elasticity of observation times to observation costs of 1/2 in a neighbourhood of $\theta = 0$. Although the functional form $\tilde{\tau}(\theta)$ is the same as in Section 3.1, it is important to notice that the resulting dynamics are very different. Because θ is Markov here, the conditional mean varies across firms, and hence

the time between observations is not equal across firms. In the previous example, the expected value of θ was constant both over time and across firms.

Point (ii) shows that as the cost becomes i.i.d. (*i.e.* $\kappa \rightarrow 0$) the optimal observation time is insensitive to the realization of the current observation cost and converges to a constant, just as in Section 3.1. Intuitively, when $\kappa \rightarrow 0$ the current observation cost is uninformative about the future observation cost, while its unconditional expectation $E[\theta]$ is the main determinant of the optimal time to the next observation and the rule becomes flat. In Figure 2, we illustrate the effects of the persistence parameter κ on the shape of the decision rule $\tau(\theta)$ at parameters of the model that imply an average frequency of price changes of about 1.3 adjustments/observations per year, and an average size of price changes equal to 10% (see the Online Appendix for more details on the parameters choice). It is apparent that the decision rule has a square root-like shape only when the persistence of the cost shocks is sufficiently high. Otherwise the decision rule is flat and unresponsive to changes in the costs.

4.3. *The scale of observation cost and the optimal observation times*

The results of Propositions 5 and 6 relate to the findings of Reis's (2006) Proposition 4, which shows that the optimal time between observations is (in our notation) approximately equal to $\tau(\theta) \approx \sqrt{2\theta/B}$ for $E[\theta] \rightarrow 0$. We have extended that characterization to a more general case when $E[\theta]$ is not necessarily close to zero, and showed how the shape of the optimal observation times depends both on the average observation cost, $E[\theta]$, and on the persistence of the observation cost, κ . The extension to non-negligible $E[\theta]$ is important: in the case of negligible costs (*i.e.* when $E[\theta] \rightarrow 0$, a situation Reis refers to as "costless planning") the mean frequency of observation is very high (diverging in the limit), so that the economy becomes one with flexible prices. A new and important result of this extension is to show that, when the costs are non-negligible, the optimal firm behaviour is well approximated by a square root formula only if the observation costs are sufficiently persistent. Notice in particular that our propositions imply that, as the realizations of the observation cost converge to zero ($\theta \downarrow 0$), the optimal observation time will not converge to zero if the persistence is low enough ($\kappa < \bar{\kappa}$), and that $\tau(\theta)$ will converge to a constant as the observation cost process converges to i.i.d. ($\kappa \rightarrow 0$).²³ Our results highlight the crucial role of $E[\theta]$ for the shape of $\tau(\theta)$: the thresholds $\underline{\kappa}$ and $\bar{\kappa}$ are increasing in $E[\theta]$. To highlight the role of the scale of observation costs, in Figure 3 we plot the optimal $\tau(\theta)$ when, everything else being equal, we scale the distribution of observation costs \hat{F} so that $E[\theta]$ is 1/1000 smaller than in the case of Figure 2. It is shown that the optimal time to the next observation is indeed well approximated by the square root expression, even at very low levels of persistence. However, the negligible observation costs used in the case of Figure 3 imply a sizeable increase in the frequency of information gathering, from about 1.3 observation per year (in the baseline case) to about 40 observations per year. In such a case, the quantitative behaviour of the model is close to flexible prices.

4.4. *The variation of costs and observation times*

We have shown that combinations of low average observation costs and/or a high persistence of the observation cost are associated with a higher elasticity of observation times with respect to the observation costs. This result is useful to determine the mapping from the variation of

23. It is immediate to see the difference with the square root approximation where the observation time converges to zero with an elasticity of one-half as the observation cost θ converges to zero.

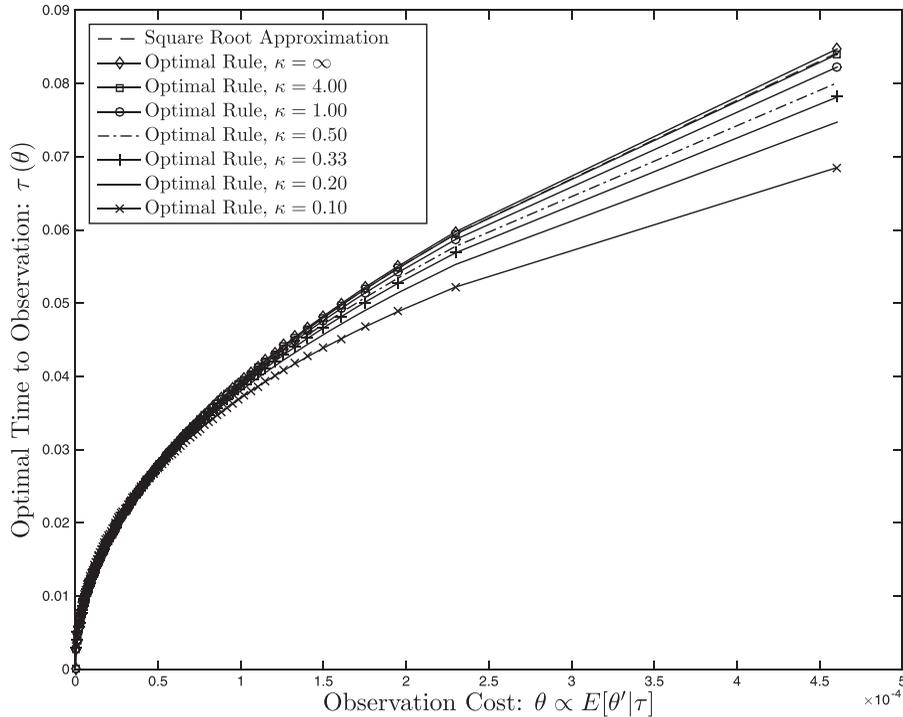


FIGURE 3

The optimal observation times as all costs gets negligible: $E[\theta] \rightarrow 0$

Note: The figure reports the optimal time to the next observation $\tau(\zeta)$, in years, as a function of the signal value ζ on the horizontal axis, and for different values of persistence, as measured by κ . The distribution \hat{F} is assumed to be exponential with average observation cost $E[\theta]=0.05/1000$: such negligible costs implies about forty observations per year on average. We set the drift $\mu - \phi + \sigma^2/2=0$ to obtain infrequent price changes, and $\sigma=0.114$ to match an average size of price changes equal to 10%. The other parameters are $\eta=5$, $\lambda=0.25$, and $\rho=0.02$. For more details see the Online Appendix.

observation costs to the variation of observation times. The next proposition uses these results to characterize the coefficient of variation of observation times (excluding substitutions) in terms of the coefficient of variation of the observation cost (see Appendix A for the proof).

Proposition 7. *Let $E(\theta) \rightarrow 0$ and the coefficient of variation of observation cost be such that $CV(\theta) > 0$ as $E(\theta) \rightarrow 0$. Let $CV_K(\tau)$ denote the coefficient of variation of the times between consecutive observations conditional on no substitutions. We have that*

$$CV_K(\tau) = \frac{1}{2} CV(\theta) + o\left(\sqrt{E(\theta)}\right). \tag{26}$$

An increase in the coefficient of variation of observation costs affects the cross-sectional variation of observation times with a coefficient of 1/2. The result follows from using that, as shown by Reis’s (2006) Proposition 4, the optimal policy is approximated by $\tilde{\tau}(\theta) \approx \sqrt{2\theta/B}$ as $E[\theta] \rightarrow 0$, so that the elasticity of the time between observations to the observation cost is one-half. However, as we showed above, the elasticity of observation times to the observation cost may be much lower in the general case when $E[\theta]$ is away from zero. To study the mapping from the variation in observation cost to the variation in observation times when $E[\theta]$ is away from

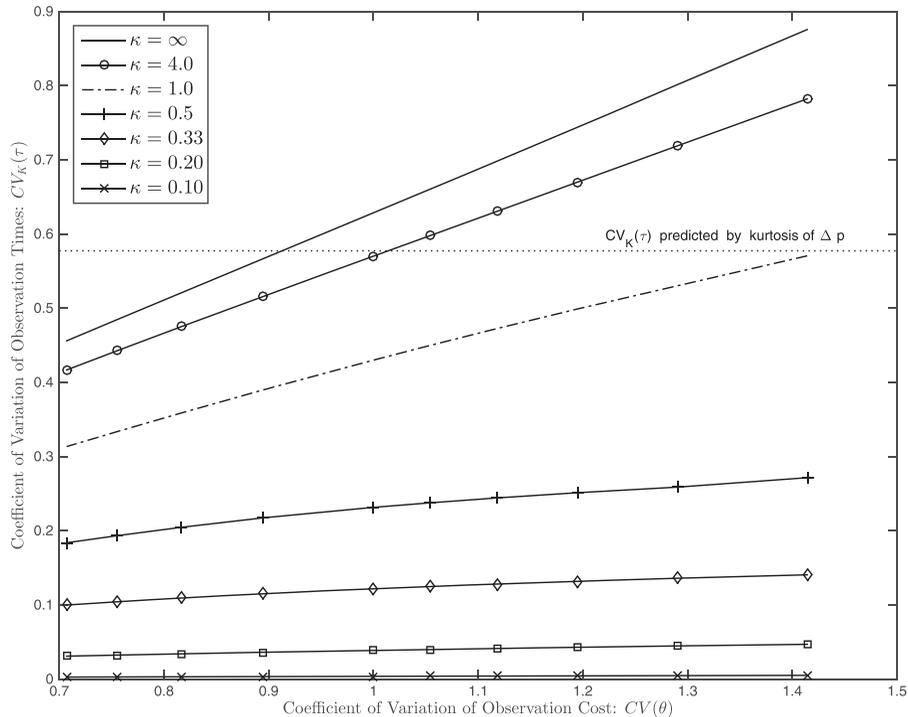


FIGURE 4

The coefficient of variation of observation times

Note: The figure reports the coefficient of variation of times between observations (in years, including substitutions) as a function of the coefficient of variation of observation cost, and for different values of persistence, as measured by κ , which is the inverse of the (yearly) arrival rate. The distribution \hat{F} is assumed to be $\text{Gamma}(\alpha_1, \alpha_2)$, where α_1 determines the coefficient of variation of observation costs upon a new draw, and α_2 is chosen to match an yearly average frequency of observations/adjustments equal to 1.3. We set the drift $\mu - \phi + \sigma^2/2 = 0$ to obtain infrequent price changes, and $\sigma = 0.114$ to match an average size of price changes equal to 10%. The other parameters are $\eta = 5$, $\lambda = 0.25$, and $\rho = 0.02$. For more details see the Online Appendix.

zero, we assume that \hat{F} is a Gamma distribution with shape parameter α_1 and scale parameter α_2 . The parameter α_1 determines the coefficient of variation of the observation cost θ . The parameter α_2 is chosen so that the average frequency of observations is 1.3 on a yearly basis. Together with the assumption of no drift in nominal marginal cost, this matches the frequency of price changes estimated by Nakamura and Steinsson (2008) in the U.S. data. We notice that targeting $1/E[\tau(\theta)] = 1.3$ implies an implicit target for $E[\theta]$, an important parameter in the analysis above to determine the elasticity of optimal observation times to observation cost.

In Figure 4, we plot the model-implied coefficient of variation of observation times ($CV_K(\tau)$ on the vertical axis) as a function of the coefficient of variation of observation costs, for different values of cost persistence, κ . The horizontal dotted line in the figure indicates the level of variation in observation times that is consistent with the kurtosis of the empirical distribution of price changes.²⁴ This shows that the model can predict variation in observation times consistent with the distribution of price changes only when persistence in the observation cost is sufficiently

24. This uses an estimated value for the kurtosis of price changes around 4, consistent with Alvarez *et al.* (2014), and the results in Proposition 3, which yields the coefficient of variation $CV_K(\tau) = 0.58$.

TABLE 1
 Cumulated output effect of a monetary shock $\delta=0.01$ (1%): $\mathcal{M}(\delta) \times 100$

	$\kappa=0.10$	$\kappa=0.33$	$\kappa=1.00$	$\kappa=4.00$	$\kappa=\infty$
$CV(\theta)=0.75$	0.38	0.39	0.43	0.46	0.48
$CV(\theta)=1.00$	0.38	0.39	0.46	0.51	0.54
$CV(\theta)=1.20$	0.38	0.39	0.48	0.56	0.60
$CV(\theta)=1.40$	0.38	0.39	0.51	0.62	0.68

Note: Values are percentages, *i.e.* $\mathcal{M}(\delta) \times 100$. The table reports the cumulated output effects to a $\delta=0.01$ (1%) monetary shock as a function of the coefficient of variation of observation cost, and for different values of persistence, as measured by κ , which is the inverse of the (yearly) arrival rate. The cumulated output response is $\mathcal{M}(\delta)$ defined in Proposition 2, with $\epsilon=1$. The distribution \hat{F} is assumed to be *Gamma*(α_1, α_2), where α_1 determines the coefficient of variation of observation costs upon a new draw, and α_2 is chosen to match an yearly average frequency of observations/adjustments equal to 1.3. We set the drift $\mu - \phi + \sigma^2/2=0$ to obtain infrequent price changes, and $\sigma=0.114$ to match an average size of price changes equal to 10%. The other parameters are $\eta=5$, $\lambda=0.25$, and $\rho=0.02$. For more details see the Online Appendix.

large. With high persistence the model predicts a sizeable variation in observation times even when the variation in observation cost is moderate.²⁵ Given the results of Proposition 2, the size of the real effect of monetary shocks can be readily computed from Figure 4. Table 1 reports the percentage cumulated output response, *i.e.* $\mathcal{M}(\delta) \times 100$, to a 1% monetary shock, *i.e.* $\delta=0.01$, for different combinations of $CV(\theta)$ and κ .²⁶ For instance, when $CV(\theta)=1$, the cumulated output effect is 0.38% if the average duration of the observation cost is one-tenth of a year (*i.e.* $\kappa=0.1$), and 0.54% if the observation cost is permanent (*i.e.* $\kappa \rightarrow \infty$).

We conclude this section by stressing that the process of optimal observation times does not inherit the properties of the process of the observation cost. In fact as the process of observation cost converges to i.i.d., *i.e.* $\kappa \rightarrow 0$, the optimal time between observations converges to a constant. The economic motivation is that when $\kappa \rightarrow 0$ the current observation cost is uninformative about the cost of the next observation, so that the firm plans according to its best estimate which is the expected value $E[\theta]$. An immediate implication of these result is that the model with a Markovian stochastic process of observation cost cannot generate i.i.d. observation times, which has been a main focus of the “rational inattentiveness” literature (see Reis, 2006, Mankiw and Reis, 2002, Carvalho and Schwartzman, 2012). We want to remark that this is not a feature of our particular specification of the distribution function F , but it is a more general result. In order for observations to be i.i.d. over time, we need an environment where the signal about the future observation costs is *both* informative *and* independent of all other shocks (including the current observation cost). This structure cannot be achieved with an environment where the observation cost follows a stochastic process that evolves independently of the observation decisions of the firm. We will therefore consider a specification of the observation cost process that delivers i.i.d. observations times in the next section.

5. OPTIMAL I.I.D. OBSERVATION TIMES

This section presents a specification of the firm problem where the optimal consecutive observation times are i.i.d. random variables. We are interested in this case mostly because a large literature has investigated the implications of i.i.d. observation times for the real effects of

25. In the case of $\kappa=\infty$, the required standard deviation of observation cost is of the same order of magnitude of its mean, *i.e.* $CV(\theta) \approx 1$.

26. In the example we assume that the elasticity of real money balances to output is equal to $\epsilon=1$.

monetary shocks, such as Mankiw and Reis (2002) or more recently Carvalho and Schwartzman (2012). The solution to the firm problem will provide a mapping from the realization of the observation costs to the optimal time between consecutive observations, τ . As in the previous section, we will use this mapping to obtain a relationship from the coefficient of variation of observation costs $CV(\theta)$ to the coefficient of variation of observation times, $CV_K(\tau)$, and hence to the size of the real effects of monetary shocks.

Optimal i.i.d. observation times cannot be produced in an environment where the observation cost follows a stochastic process that evolves independently of the observation decisions of the firm, as in the previous section. In fact, persistence in the opportunity cost of observations leads to predictability and correlation of observation times.²⁷ To generate i.i.d. observation times, we have to assume that, upon an observation, the firm receives a signal ζ about the future observation cost θ' , that is independent of the current observation cost, θ (see Figure 1 for an illustration of the timeline of θ and ζ). Formally, the observation times are i.i.d. over time if and only if $G(\zeta|\theta_0) = G(\zeta|\theta_1)$ for all ζ and all pairs θ_1, θ_0 , and hence $G(\zeta|\theta) \equiv \hat{G}(\zeta)$ for all ζ and θ .

The observation cost θ' will be drawn from the distribution $F(\theta', \tau|\zeta)$, that depends both on the realization of the signal ζ and on the time τ elapsed since then. A simple tractable case is one where such distribution is

$$F(\theta', \tau|\zeta) = e^{-\gamma\tau} \mathbf{1}_{\theta' \geq \zeta} + (1 - e^{-\gamma\tau}) \hat{F}(\theta'). \quad (27)$$

The parameter γ determines the information content of the signal, which is perfectly precise if $\gamma = 0$, and has no information as $\gamma \rightarrow \infty$. Notice that a higher value of the signal ζ is associated with a higher expected observation cost: $E[\theta'|\zeta, \tau] = e^{-\gamma\tau} \zeta + (1 - e^{-\gamma\tau}) \int \theta' d\hat{F}(\theta')$. In this case, the realization of the signal coincides with the cost of the next observation with a probability that decreases as time elapses, otherwise the future observation cost will be drawn from the distribution \hat{F} (independent of ζ). This specification of F captures the idea that, as time passes, the precision of the information about the future observation cost depreciates, while preserving tractability.

Given equation (27) and the assumption $G(\zeta|\theta) \equiv \hat{G}(\zeta)$ for all ζ and θ , we can rewrite the firm problem in equation (20) as:

$$v(\zeta) = \max_{\tau \in \mathbb{R}_+} \int_0^\tau e^{-(\rho+\lambda-a)t} dt + e^{-(\rho+\lambda-b)\tau} (E[v] - (e^{-\gamma\tau} \zeta + (1 - e^{-\gamma\tau}) E[\theta])), \quad (28)$$

where $E[v] \equiv \int v(\zeta') d\hat{G}(\zeta')$ and $E[\theta] \equiv \int \theta' d\hat{F}(\theta')$.²⁸ The proposition below characterizes the decision rule in this case (see Appendix A for the proof). We will use such characterization to compare to the decision rule obtained in the case of persistent observation cost in the previous section.

Proposition 8. *Assume that $F(\cdot)$ is given by equation (27) with $0 < \gamma < \infty$ and that $v > E[\theta]$. Then: (i) there exists a ζ^* so that $0 < \tau(\zeta) < \infty$ and $\tau'(\zeta) > 0$ for all $\zeta^* \leq \zeta < \infty$; moreover if $\zeta^* > \underline{\zeta}$, then $\tau(\zeta) = 0$ for all $\underline{\zeta} \leq \zeta < \zeta^*$; (ii) if $\underline{\zeta} \geq 0$, then $\tau(\underline{\zeta}) > 0$ and $\zeta^* = \underline{\zeta}$.*

Point (i) of the proposition states that the time between observations, τ , increases with the signal about the future observation cost, ζ , and does so strictly for values above a threshold ζ^* .

27. A similar result is obtained in the Online Appendix when it is the benefit of observations that varies over time.

28. Note that the particular specification of the distribution \hat{F} is immaterial for this problem, as only the associated mean $E[\theta]$ matters.

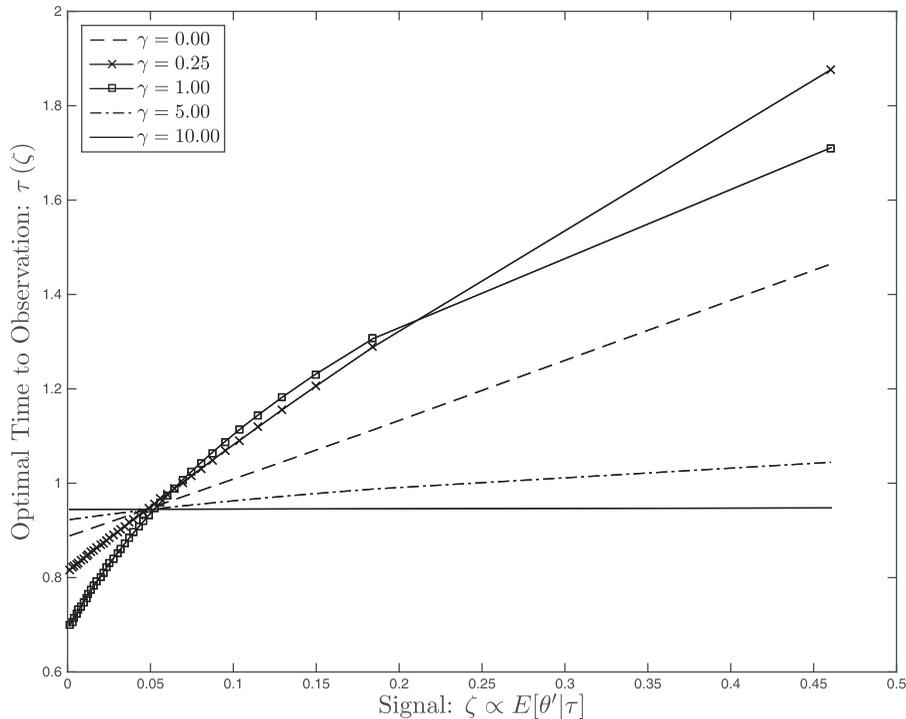


FIGURE 5

Optimal i.i.d. observation times as a function of the signal on the observation cost

Note: The figure plots the optimal time to the next observation $\tau(\zeta)$, in years, as a function of the signal value ζ on the horizontal axis, and for different values of precision, as measured by $1/\gamma$. The distribution \hat{G} is assumed to be exponential with average $E[\zeta]=0.05$, and we further assume $E[\theta]=E[\zeta]$. We set the drift $\mu - \phi + \sigma^2/2=0$ to obtain infrequent price changes, and $\sigma=0.114$ to match an average size of price changes equal to 10%. The other parameters are $\eta=5$, $\lambda=0.25$, and $\rho=0.02$. For more details see the Online Appendix.

Intuitively, a higher realization of ζ is associated with a higher expected observation cost, so the firm delays the time of the next observation. When the realization of such signal is low enough, *i.e.* $\zeta < \zeta^*$, the expectation of the next observation cost is so low that the firm finds it optimal to observe immediately. Point (ii) establishes that ζ^* must be negative for the firm to exercise the observation immediately. In fact, if the lower bound of the support of ζ is positive, *i.e.* $\zeta \geq 0$, then $\tau(\zeta) > 0$. For instance, even if $\zeta=0$ so that the cost of observing immediately would be zero, then $\tau(0) > 0$. We notice that this is true for all values of γ . As an illustration of the properties of the optimal firm policy, Figure 5 plots the optimal time to the next observation as a function of the expected cost of the next observation, for different values of γ .

Comparing Figure 5 to Figure 2 one can appreciate the different properties of the optimal time to the next observation in the cases of i.i.d. and Markov observation times, respectively. These differences are particularly evident for expected observation cost arbitrarily close to zero. The economics of $\tau(0) > 0$ is very different from the case of persistent observation cost in the previous section. Here, the firm waits a strictly positive period of time before making another observation even if the cost of observing (again) immediately is zero because of an option value argument: if the observation occurs immediately the opportunity to potentially exercise an observation at a low cost would be wasted because there is no correlation between consecutive observation costs. Due to the nature of i.i.d. observation times, the setup of the problem is such that each observation

is associated with an i.i.d. draw for the distribution of the observation cost to be paid upon the next observation. A firm that draws a signal of low (future) observation cost will wait for the best moment to use this option. The option value of waiting will turn out to be relevant to determine the smallest time between observations. For instance, one important implication is that, to generate a positive probability of arbitrarily small durations of the time between observations—*e.g.* in an exponential distribution—the distribution of observation costs needs to have positive mass on negative costs, *i.e.* $\underline{\theta} < 0$ and $\int_{\underline{\theta}}^0 d\hat{G}(x) > 0$.²⁹ This result is interesting as a large part of the literature that developed after Mankiw and Reis (2002) has focused on quantifying the predictions of models of rational inattentiveness in which the distribution of the times between observations indeed features positive mass on arbitrarily small durations. We will come back to this issue in Section 5.1. More interestingly, the option value of waiting also reduces the variability of observation times that follows from the variability of observation cost. This in turn reduces the ability of such model to produce large variation in observation times and, therefore, large real effects of monetary shocks. We discuss this in more detail later.

Next we consider two special cases for the informativeness of the signal, for which we have an (almost) closed form solution for the optimal policy. In the case where $\gamma = \infty$, the signal carries no information about the observation cost to be paid, and hence the model coincides with the case characterized in Proposition 4, where the policy $\tau(\cdot)$ does not depend on the realization of ζ . At the other extreme, $\gamma = 0$, the signal is perfectly informative about the cost to be paid at the next observation. The next two propositions describe the optimal policy and the mapping from the variability of observation cost to the variability of observation times in the $\gamma = 0$ case.

Proposition 9. *Let $F(\theta', \tau|\zeta) = \mathbf{1}_{\theta' \geq \zeta}$ for all ζ , *i.e.* $\gamma = 0$ in equation (27). The optimal policy $\tau(\zeta)$ is given by*

$$\tau(\zeta) = \begin{cases} 0 & \text{if } \underline{\zeta} \leq \zeta \leq E[v] - \bar{v} \\ \frac{1}{B} [\ln(\bar{v}) - \ln(E[v] - \zeta)] & \text{if } E[v] - \bar{v} < \zeta < E[v] \\ \infty & \text{if } \zeta \geq E[v] \end{cases}, \quad (29)$$

where the expected value $E[v] = \int_{\underline{\zeta}}^{\infty} v(\zeta') d\hat{G}(\zeta')$ is the solution of an explicit equation that is displayed in the proof. If $\underline{\zeta} \geq 0$, then: (i) $E[v] < \bar{v}$ and $\tau(\zeta) > 0$ for all $\zeta \geq \underline{\zeta}$; (ii) if \hat{G} is first-order stochastically higher, then $E[v]$ is smaller; and (iii) if \hat{G} is second order stochastically higher, then $E[v]$ is higher.

See Appendix A for the proof. When $\gamma = 0$ the optimal policy $\tau(\zeta)$ is known in closed form up to the value of the expected value function $E[v]$. It features all the properties of the case of imperfectly informative signals ($\gamma > 0$), with the exception of being characterized by an asymptote, so that if $\zeta \geq E[v]$, then $\tau = +\infty$. This is because with $\gamma = 0$ the expected cost of the next observation does not revert to its unconditional mean as time elapses, so that if it is large enough ($\zeta \geq E[v]$) making an observation will never be optimal for the firm. Points (ii) and (iii) of the proposition highlight how the properties of the distribution of signals, which coincide with observation costs in this case, affect the value of the firm.³⁰

29. Notice that a negative observation cost is not allowed in the model where the costs are Markovian because that would give the firm the possibility of unbounded profits by simply collecting the fees infinitely many times as long as the cost remains negative.

30. We notice that in this case the value of $E[\theta]$ is immaterial as observation costs are never drawn from the distribution \hat{F} .

The expression in equation (29) is particularly useful to study the elasticity of $\tau(\zeta)$ with respect to ζ . In particular, we note that $\tau'(\zeta) = 1/[B(E[v] - \zeta)]$ for $\zeta \in (E[v] - \bar{v}, E[v])$. This implies that when the benefit of information B and/or the continuation value $E[v]$ are larger, then the slope of the optimal policy with respect to the observation cost is smaller. This suggests that economies with a higher frequency of i.i.d. observations are characterized by a lower sensitivity of observation times to variation in observation costs.³¹ The next proposition uses the analytical results of Proposition 9 to characterize the coefficient of variation of observation times (excluding substitutions) in terms of the coefficient of variation of the observation cost in the case of i.i.d. observations with perfect predicability (see Appendix A for the proof).

Proposition 10. *Let $F(\theta', \tau|\zeta) = \mathbf{1}_{\theta' \geq \zeta}$ for all ζ , i.e. $\gamma = 0$ in equation (27), and $G(\zeta|\theta) = \hat{G}(\zeta)$ for all ζ and θ . Moreover, let $E(\zeta) = \int_{\underline{\zeta}}^{\infty} \zeta d\hat{G}(\zeta)$ and the coefficient of variation of observation cost be such that $CV(\zeta) > 0$ as $E(\zeta) \rightarrow 0$. Let $CV_K(\tau)$ denote the coefficient of variation of times between observations conditional on no substitutions. We have that*

$$CV_K(\tau) = \frac{1}{2}(\rho + \lambda - b)\bar{\tau} CV(\zeta) + o\left(\sqrt{E(\zeta)}\right), \tag{30}$$

where $\bar{\tau} = \sqrt{2E(\zeta)/(b-a)}$ is a measure of the mean time between observations.

The proposition shows that the mapping from $CV(\zeta)$ to $CV_K(\tau)$ depends on two terms. The first term, i.e. $(\rho + \lambda - b)$, is related to the discount rate in equation (28) and typically has the order of magnitude of the interest rate. The second term, i.e. $\bar{\tau} = \sqrt{2E(\zeta)/(b-a)}$, is of the same order of magnitude of the average duration of observation times. For instance, an economy with a yearly discount rate of 5% and an average time between observations of 1 year, would be characterized by $CV_K(\tau)/CV(\zeta) \approx 0.025$. Such a small slope of the variability in observation times with respect to observation costs is explained by the option value argument discussed above.

To study the mapping from the variation in observation cost to the variation in observation times in the more general case when signals are not perfectly informative $\gamma > 0$, we do the following exercise. We assume that the distribution \hat{G} is a Gamma distribution with shape parameter α_1 and scale parameter α_2 . The parameter α_1 determines the coefficient of variation of the observation cost θ , while the parameter α_2 is chosen so that the average frequency of observations is 1.3 on a yearly basis which, together with the assumption of no drift in nominal marginal cost, matches the estimated frequency of price changes in the U.S. data. In Figure 6, we plot the model-implied coefficient of variation of observation times ($CV_K(\tau)$) on the vertical axis as a function of the coefficient of variation of observation costs, for different values of γ . The figure also plots the level of variation in observation times that is consistent with the kurtosis of the distribution of price change, as explained in Section 4. The results clearly show that a much higher variability in observation costs is needed to produce a given variability in observation times compared to the case with persistent shocks (see Figure 4 for comparison), for all values of γ . We conclude that the model of i.i.d. observations has a hard time generating large real effects of monetary shocks: for the parametrization presented in Figure 6 there is no level of $CV(\zeta)$, no matter how large, that is able to produce the variation of $CV_K(\tau)$ consistent with the kurtosis of the distribution of price changes.

31. Note that both the values of B and $E[v]$ are positively associated with the average frequency of observations: as B increases, observing more frequently is optimal; if the average frequency of observations is optimally higher, then $E[v]$ increases and gets closer to its upper bound \bar{v} .

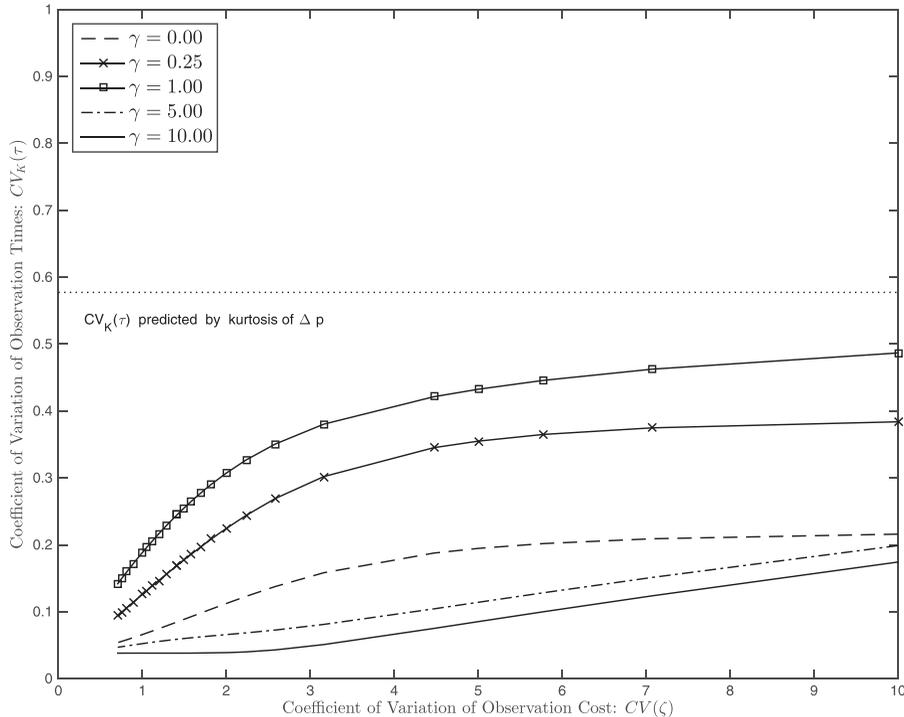


FIGURE 6

Coefficient of variation of observation times with i.i.d. observation cost

Note: The figure reports the coefficient of variation of times between observations (including substitutions) as a function of the coefficient of variation of signals about future observation cost, ζ , and for different values of precision, as measured by $1/\gamma$. The distribution \hat{F} is assumed to be $Gamma(\alpha_1, \alpha_2)$, where α_1 determines the coefficient of variation of observation costs upon a new draw, and α_2 is chosen to match an yearly average frequency of observations/adjustments equal to 1.3. We set $E[\theta] = E[\zeta]$. We set the drift $\mu - \phi + \sigma^2/2 = 0$ to obtain infrequent price changes, and $\sigma = 0.114$ to match an average size of price changes equal to 10%. The other parameters are $\eta = 5$, $\lambda = 0.25$, and $\rho = 0.02$. For more details see the Online Appendix.

5.1. An application: microfounding exponential i.i.d. observations

In this section, we use our results to solve a “reverse engineering problem”: for given distribution of times between consecutive observations K as defined in equation (4), we find parameters for the firm problem, including the distribution of the observation cost, so that aggregating the resulting optimal decision rules $H(t'|t)$ we obtain the target distribution K . While we could do this for any K , a particular interesting application, given the results by Mankiw and Reis (2002), is to use an i.i.d. exponential K . The aim of this is to identify the type of microfoundation and the parameters of the firm problem in Section 5 that would be consistent with this popular framework.

Let \hat{K} denote the invariant distribution of times between consecutive observations conditional on no substitution. We note that the invariant distribution of observation times K including substitutions (which are exponentially distributed with parameter λ by assumption) is exponential if and only if \hat{K} is exponential, and satisfies $K(t) = e^{-\lambda t} \hat{K}(t)$ for each $t \geq 0$. The next proposition describes the distribution of observation cost \hat{G} that microfounds an exponential distribution of observation times, where $\hat{K}(t) = \exp(-\xi t)$ for some $\xi > 0$, in the case where observations are i.i.d. and signals are perfectly informative, *i.e.* $\gamma = 0$.

Proposition 11. Consider the case when observation times are i.i.d. with $\gamma = 0$. Let the invariant exponential distribution of times between consecutive observations (excluding substitutions) be $\hat{K}(t) = \exp(-\xi t)$ with parameter $\xi > 0$. Then the density of the observation costs (and signals) is a displaced beta distribution $Beta(\alpha_l, \alpha_r)$ with left parameter $\alpha_l = 1$, right parameter $\alpha_r = \xi/B$, and support $(\underline{\zeta}, \underline{\zeta} + \bar{v})$, so that the density $\hat{g}(\cdot) = \hat{G}'(\cdot)$ is given by:

$$\hat{g}(\zeta) = \frac{\xi}{\bar{v}B} \left(\frac{E[v] - \zeta}{\bar{v}} \right)^{\frac{\xi}{B} - 1} \text{ for } \zeta \in (\underline{\zeta}, \underline{\zeta} + \bar{v}), \tag{31}$$

where the expected value function and the lower bound of the distribution of cost are

$$E[v] = \underline{v} + (\bar{v} - \underline{v}) \frac{\xi \underline{v}}{1 + \xi \underline{v}}, \quad \underline{\zeta} = E[v] - \bar{v}. \tag{32}$$

The implied fraction of negative cost $\zeta \leq 0$ and the coefficient of variation of cost are

$$\hat{G}(0) = 1 - \left(\frac{\underline{v}}{\bar{v}} + \left(1 - \frac{\underline{v}}{\bar{v}}\right) \frac{\xi \underline{v}}{1 + \xi \underline{v}} \right)^{\frac{\xi}{B}} > 0, \text{ and } CV(\zeta) = \sqrt{\frac{\xi/B}{2 + \xi/B} \left(\frac{1 + \xi \underline{v}}{\underline{v}/\bar{v}} \right)}. \tag{33}$$

See Appendix A for the proof. This proposition shows that if observation times are exponential, the observation costs have to be distributed as a displaced beta distribution. The shape of this distribution is determined by $\xi/B - 1$, which depends on the ratio of the exponential parameter ξ to the benefits of information: $B = \eta(\eta - 1)\sigma^2/2$. The support of the cost is the interval $(E[v] - \bar{v}, E[v])$. Since the exponential distribution has positive probability for arbitrarily small values of observation times, the option value argument of Proposition 8 implies that the lower bound is negative. The expression for the fraction of negative cost $\hat{G}(0)$ allows us to quantify the importance of this feature.

We next comment on the comparative static of the change in the distribution of the observation cost as we vary the parameter of the exponential distribution of observation times. Note that if $\xi \rightarrow 0$ observations are very infrequent, then $\underline{\zeta} \rightarrow 0$, so that $\hat{G}(0) \rightarrow 0$ and $CV(\zeta) \rightarrow 0$. In words, to have almost no observations, the distribution of the observation cost must be almost degenerate, concentrated around its upper bound $\bar{\zeta} = \underline{v}$. At the other extreme, if $\xi \rightarrow \infty$ observations are very frequent, then $\underline{\zeta} \rightarrow \bar{v}$, the fraction $\hat{G}(0) \rightarrow 1 - \exp(-1) \approx 0.63$ and $CV(\zeta) \rightarrow +\infty$. The general case moves monotonically between these two extremes, with both the fraction of negative costs, $\hat{G}(0)$, and the coefficient of variation, $CV(\zeta)$, being increasing functions of the frequency of observations ξ .

In the left panel of Figure 7, we plot the fraction of negative costs as a function of the average frequency of observations ξ , for selected parameter values. This figure plots three lines, corresponding to three values of η , which change the value of acquiring information captured by B . Note that for $\xi \approx 1.3$, which roughly corresponds to the average frequency of price changes per year in the U.S., then $\hat{G}(0) \approx 0.55$, so that more than half of the observation costs are negative. We interpret this as saying that the option value effect is very large for this model to be quantitatively realistic, since negative observation cost have no direct economic interpretation. We also note that the value of η has marginal impact on the value of $\hat{G}(0)$. Additionally, in the right panel of Figure 7, we plot the coefficient of variation of the observation cost as a function of the frequency of observations, ξ , and three values of η . For $\xi \approx 1.3$, given our parameter values, the coefficient of variation is about 10 at $\eta = 5$. We recall that we are measuring the observation cost ($\zeta = \theta$) as a multiple of the frictionless yearly profit, implying that a large coefficient of variation can be

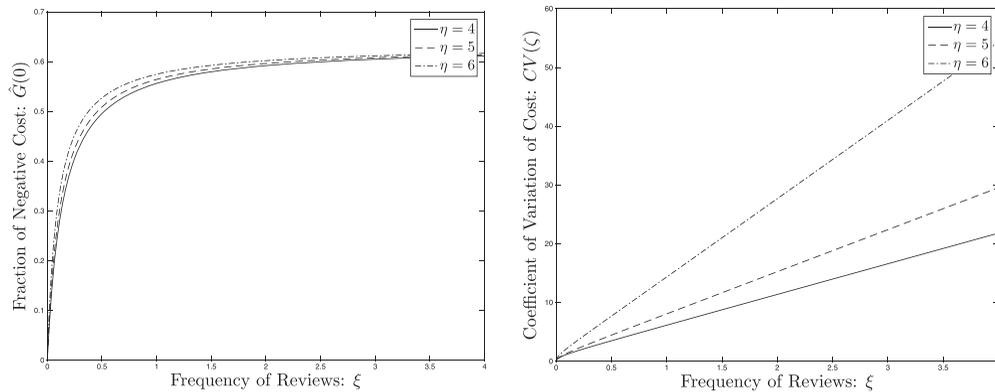


FIGURE 7

Exponential i.i.d. observation times: properties of \hat{G} as a function of ξ and η

Note: The figure reports the fraction of negative cost (left panel) and the coefficient of variation of cost (right panel) as a function of the frequency of observation on the horizontal axis, for three different values of $\eta = \{4, 5, 6\}$, associated with the distribution of Proposition 11. We set $E[\theta] = E[\xi]$. We set the drift $\mu - \phi + \sigma^2/2 = 0$ to obtain infrequent price changes, and $\sigma = 0.114$ to match an average size of price changes equal to 10% when the frequency of observations/price adjustments is 1.3 on a yearly base. The discount rate is $\rho = 0.02$. For more details see the Online Appendix.

associated with, in our view, unreasonably high values of the cost of an observation in units of profits. Finally, the coefficient of variation of cost increases with the value of η . Intuitively, as the value of information increases firms have incentives to observe sooner for a given realization of the observation cost. And so more variation in costs is needed to generate large enough durations of observations such that the observed outcomes are consistent with an exponential distribution.

6. CONCLUDING REMARKS

We explored the microfoundations, as well as the aggregation, of models where the optimal price-setting decision of the firm is subject to information gathering costs, as in the “rational inattentiveness” literature. Our analytical results unveil a few shortcomings of the current literature, as well as potentially fruitful avenues for future research. We showed that a natural modelling of the firm’s decisions, one that assumes independence of the cost process from the firm’s decisions to gather information, produces optimal observation times that are persistent and, for reasonable parameterizations, infrequent. These results, which deviate from existing models mainly in the fact that observation times are not i.i.d., are also useful for generating larger real effects of monetary policy.

Our theoretical characterization relies on a number of simplifying assumptions, whose importance seems of interest for future research. A limitation of our framework is the absence of physical costs of price adjustments, implying that prices change continuously in presence of a drift in nominal marginal cost. In Alvarez *et al.* (2011, 2015) we explore the optimal pricing decision of firms, and its aggregate implications, when firms face non-random observation and adjustment costs. An important result of these papers is that *price plans* are not optimal for reasonable parameterizations of the adjustment cost, so that prices change infrequently even in presence of a drift in marginal cost. A large literature (*e.g.* Mankiw and Reis, 2002) has emphasized the role of price plans for the propagation of monetary shocks to inflation and output. As in this literature, we allow firms to use price plans at no cost. However, the monetary shock we study causes

no predictable dynamics in nominal marginal cost so that price plans are not used by firms in response to it.

Our results on the aggregation of firms' actions and propagation of monetary shocks rely, for tractability, on a framework where there are no strategic complementarities in price setting, *i.e.* the profit-maximizing price of a firm is independent of the other firms' actions. While allowing for these strategic complementarities may increase the size of the real effects of monetary shocks, it is not easy to predict how they would affect the mapping from the cross-sectional distribution of observation times to the size of the real effects of monetary shocks.

We have focused on a model where the cost/benefit of making an observation varies because of random variation in the observation cost for simplicity and ease of comparison with the previous literature. However, there are other interesting sources of variation in the the cost/benefit ratio. For instance, relaxing the assumption that the cost of making an observation scales with firms' profits can deliver variability and persistence in observation times arising from variability and persistence of the productivity process. Alternatively, in the Online Appendix, we explore a model where observation costs are constant but their benefit varies because of time-varying volatility of the state. Finally, adding a separate cost of adjusting prices (*i.e.* a menu cost) would cause firms to differ in the gap between the price they post and the optimal one. In this specification, variation in observation times may arise from variation across firms in the benefit, as opposed to the cost, of the next observation. Further investigating and measuring alternative sources of variation in the cost/benefit of observations is a useful avenue for future research.

APPENDIX

A. PROOFS

Proof. (of Proposition 1) Consider the case where $H(\cdot|s)$ is absolutely continuous for all s . We use a guess and verify strategy, so we substitute equation (7) into equation (2) and write

$$RK(T) = -R \int_0^\infty H(T|s) k(s) ds \quad \text{for all } T \geq 0. \tag{34}$$

Note that this expression holds for $T=0$ since $K(0) = H(0|s) = 1$ for all s since both K and $H(\cdot|s)$ are right CDFs. Next, differentiate both sides of equation (34) to obtain:

$$k(T) = \int_0^\infty h(T|s) k(s) ds \quad \text{for all } T \geq 0, \tag{35}$$

which is exactly the definition of an invariant distribution given in equation (1). Consider now the case where H has positive mass on countably many values, note that for each T we can write:

$$\begin{aligned} q(T) &= R \sum_{j=1}^\infty H(T|t_j) k(t_j) = R \sum_{j=1}^\infty \sum_{\{i:t_i \geq T\}} h_{ij} k(t_j), \\ &= R \sum_{\{i:t_i \geq T\}} \sum_{j=1}^\infty h_{ij} k(t_j) = R \sum_{\{i:t_i \geq T\}} k(t_i) = RK(T), \end{aligned}$$

where the first equality uses that times between consecutive observations take countably-many values, the second uses the definition of h , the third permutes the summations, and the last one uses the definition of an invariant distribution. Note that the density of the time until the next observation is piece-wise linear, with downward jumps at the observation times $T = t_i$ for all $i \geq 1$. \parallel

Proof. (of Proposition 2) First we show that

$$\mathcal{M}(\delta) \frac{\epsilon}{\delta} \equiv \int_0^\infty Q(t) dt = R \int_0^\infty \int_t^\infty K(s) ds dt = R \int_0^\infty s K(s) ds$$

We write $A(J) \equiv \int_0^J \int_t^J K(s) ds dt$ and $B(J) \equiv \int_0^J s K(s) ds$ for all $J \geq 0$. Note $A(0) = B(0) = 0$. Secondly, using a prime to denote derivative, $B'(J) = JK(J)$ and $A'(J) = \int_0^J K(J) dt = JK(J)$, hence $A(J) = B(J)$ and taking $J \rightarrow \infty$ we obtain the desired result.

Secondly, let the unconditional variance of τ be

$$\text{Var}(\tau) \equiv - \int_0^\infty s^2 K'(s) ds - \left(- \int_0^\infty s K'(s) ds \right)^2 = 2 \int_0^\infty s K(s) ds - \left(\int_0^\infty K(s) ds \right)^2,$$

where the second equality comes from integrating by parts. The latter, together with the definition of R , i.e. $\int_0^\infty K(s) ds = 1/R = E[\tau]$ implies that $\int_0^\infty s K(s) ds = 1/2 \text{Var}(\tau) + (E[\tau])^2$, which immediately implies the result. \parallel

Proof. (of Proposition 3) Consider first the distribution of log-price changes of a firm observing/adjusting τ periods after the last observation/adjustment, $\Delta \log(p(\tau))$. Conditional on the next observation taking place in τ periods, the distribution of log-price changes upon the next observation/adjustment is normal with mean and variance equal to 0 and $\sigma^2 \tau$, respectively. Let index the distribution of $\Delta \log(p(\tau))$ by τ . Recall that the kurtosis of a random variable x with zero mean is defined as $Kurt(x) = m_4(x)/(m_2(x))^2$, where $m_s(x) = \mathbb{E}[x - \mathbb{E}(x)]^s$ is the s^{th} centred moment of x . Therefore, the second and fourth centred moments of $\Delta \log(p(\tau))$ are equal to $m_2(\tau) = \sigma^2 \tau$ and $m_4(\tau) = 3 \sigma^4 \tau^2$, respectively. Consider now the cross-sectional distribution of price changes, $\Delta \log(p)$. Such distribution is given by the mixture of normals arising from the different firms drawing different times to the next observation, τ , each with density $k(\tau)$.

The second moment of such mixture of normals is:

$$\int m_2(\tau) k(\tau) d\tau = \sigma^2 \int \tau k(\tau) d\tau.$$

The fourth moment of such mixture of normals is given by the weighted average of the fourth moments of each normal, $m_4(\tau)$:

$$\int m_4(\tau) k(\tau) d\tau = 3 \sigma^4 \int \tau^2 k(\tau) d\tau.$$

Thus, the kurtosis of the distribution of log-price changes is given by:

$$Kurt(\Delta \log(p)) = \frac{3 \left(\int \tau^2 k(\tau) d\tau \right)}{\left(\int \tau k(\tau) d\tau \right)^2} = \frac{3 \left(\text{Var}(\tau) + (E[\tau])^2 \right)}{(E[\tau])^2} = 3 \left((CV(\tau))^2 + 1 \right). \parallel$$

Proof. (of Proposition 4) Taking the first-order condition with respect to τ in the right-hand side of the Bellman equation $\hat{\tau} = \frac{1}{B} [\ln(\hat{v}) - \ln(\hat{v} - E[\theta])]$. Using this to eliminate τ from the Bellman equation we obtain

$$\hat{v} = \begin{cases} \bar{v} & \text{if } E[\theta] = 0 \\ \text{smallest solution of } \hat{v} = \underline{v} + B \underline{v} \hat{v} \left(\frac{\hat{v} - E[\theta]}{\hat{v}} \right)^{\frac{1}{B\underline{v}}} & \text{if } 0 < E[\theta] < \underline{v} \\ \underline{v} & \text{if } E[\theta] \geq \underline{v} \end{cases}, \quad (36)$$

which can be written as: $x = s(x; \hat{\theta}, \alpha) \equiv \underline{v}/\hat{v} + \alpha \left(x - \hat{\theta} \right)^\alpha$, where $x \equiv \hat{v}/\underline{v}$ and $\alpha = B\underline{v} < 1$ under our maintained assumptions. We are looking for a solution with $\underline{v}/\hat{v} \leq x \leq 1$. Note that for $\hat{\theta} = 0$ then there is a unique solution $\hat{v} = \bar{v}$ –to see this note that $s(x; \hat{\theta}, \alpha)$ is strictly convex in x and that $s'(1, 0, \alpha) = 1$. For $\hat{\theta} > 0$ there is also a solution of this equation with $x > 1$, which does not correspond to a solution for the value function. For values of $\hat{\theta}$ where there are two solutions, the smallest one is in the desired range. Since increases in $\hat{\theta}$ shift horizontally the function s then the smallest solution decreases with $\hat{\theta}$. Since increases in α shift the function s up, then the smallest solution increases with α . Finally, if $\hat{\theta} = \underline{v}/\hat{v}$ the smallest solution is $\hat{v} = \underline{v}$. If $\hat{\theta} > \underline{v}/\hat{v}$ then there is only one solution of $x = s(x)$ which is not the solution of the value function, and hence in this case $\hat{\tau} = +\infty$ and $\hat{v} = \underline{v}$. \parallel

Proof. (of Proposition 5) We first prove part (i). We start by arguing that with $\kappa > 0$, then $\tau(\theta) > 0$ for all $\theta > 0$. This follows because if $\tau(\theta) = 0$ with $\theta > 0$ then the expected time until there is a change is strictly positive, and at each time there is a strictly positive cost, so the expected discounted cost diverges to $+\infty$.

Next we derive the first-order condition to the firm problem and show it is sufficient for an optimum. For this, note that the derivative of the objective function in the Bellman equation (25) with respect to τ is equal to $e^{-(\rho+\lambda-\theta)\tau} [M(\tau) - N(\tau, \theta)]$ where these functions are given by:

$$M(\tau) \equiv 1 - (\rho + \lambda - b) e^{B\tau} (E[v] - E[\theta]) \quad \text{and}$$

$$N(\tau, \theta) \equiv (\rho + \lambda - b + 1/\kappa) e^{(B-1/\kappa)\tau} [v(\theta) - \theta - (E[v] - E[\theta])]. \quad (37)$$

First, we note several properties of M and N at $\theta > 0$: (a) for (i) to hold, i.e. $\tau(\theta) > 0$ for any $\theta > 0$, it must be the case that $M(0) - N(0, \theta) > 0$; (b) $M(\tau) - N(\tau, \theta)$ is continuous in τ and $\lim_{\tau \rightarrow \infty} M(\tau) - N(\tau, \theta) = -\infty$ for all $\theta < \infty$. From (a) and (b) we obtain that there is at least one finite solution to the first-order condition, $\tau(\theta) < \infty$, for all $\theta < \infty$. Next, we show that there is a unique local maximum. Differentiating the first-order condition with respect to τ we have

$S(\tau) \equiv M_\tau(\tau) - N_\tau(\tau, \theta) = -B + \kappa^{-1} (1 - (\rho + \lambda - b)e^{B\tau} (E[v] - E[\theta]))$. We prove now that the objective function has at least one local maximum where the first-order condition holds. Let us denote by τ_1 the smallest local maximum. Notice that $S(\tau_1) < 0$ by definition of a local maximum. If there would be another local maximum $\tau_2 > \tau_1$, then there must be a value of $\tau_m \in (\tau_1, \tau_2)$ that is a local minimum requiring $S(\tau_m) \geq 0$. But notice that the function $S(\tau)$ is strictly decreasing in τ which is a contradiction of $S(\tau_m) \geq 0 > S(\tau_1)$ as $\tau_m > \tau_1$. Finally, notice that as $\theta \rightarrow \infty$, then $\tau(\theta) \rightarrow \infty$. This follows from equation (37) given that $\theta \rightarrow \infty$ implies that $N(\tau, \theta) \rightarrow -\infty$ and the first-order condition can be satisfied only if $\tau \rightarrow +\infty$ so that also $M(\tau)$ diverges to $-\infty$.

Next we use the first-order condition to prove $\tau'(\theta) > 0$ if $\theta > 0$ and $\lim_{\theta \rightarrow \infty} \tau'(\theta) = 0$. The implicit function theorem gives

$$\frac{\partial \tau}{\partial \theta} = \frac{N_\theta}{M_\tau - N_\tau} > 0,$$

where $N_\theta = (\rho + \lambda - b + 1/\kappa)e^{(B-1/\kappa)\tau} (v'(\theta) - 1) < 0$ because $v'(\cdot) \leq 0$, and $M_\tau - N_\tau < 0$ because $\tau(\theta)$ is a maximum. Finally, $\lim_{\theta \rightarrow \infty} \tau'(\theta) = 0$ follows because $M_\tau - N_\tau$ diverges to $-\infty$ while N_θ either converges to zero or diverges to $-\infty$ but at a lower rate than $M_\tau - N_\tau$ as θ converges to zero.

Next, we prove part (iii). We argue that if $1/\kappa$ is low enough then $\lim_{\theta \rightarrow 0^+} \tau(\theta) = 0$. We notice that if $\tau = 0$ solves the first-order condition to the firm problem at $\theta = 0$ then the firm value is given by $v(0) = \frac{1+1/\kappa(E[v]-E[\theta])}{\rho+\lambda-b+1/\kappa}$; substituting this expression for $v(0)$ into equation (37) we obtain that indeed $\tau(0) = 0$ is an extreme point. We are left to show that $\tau(0) = 0$ is a maximum which requires the second derivative at $\tau = 0$ to be negative, $S(0) = -B + \kappa^{-1} (1 - (\rho + \lambda - b)(E[v] - E[\theta])) < 0$. By using $E[v] > 1/(\rho + \lambda - a)$, a sufficient condition for the latter is $-B + \kappa^{-1} [1 - (\rho + \lambda - b)(1/(\rho + \lambda - a) - E[\theta])] < 0$, implying $\kappa > \bar{\kappa} \equiv \frac{B+(\rho+\lambda-a)(\rho+\lambda-b)E[\theta]}{B(\rho+\lambda-a)}$. Finally, if $\lim_{\theta \rightarrow 0^+} \tau(\theta) = 0$ then $\lim_{\theta \rightarrow 0^+} \tau'(\theta) = +\infty$ follows because $N_\theta(\tau; \theta) = (\rho + \lambda - b + 1/\kappa)e^{(b-a-1/\kappa)\tau} (v'(\theta) - 1)$ diverges as $\theta \rightarrow 0^+$ because $v'(\theta) = \frac{-e^{-(\rho+\lambda-b+1/\kappa)\tau(\theta)}}{1-e^{-(\rho+\lambda-b+1/\kappa)\tau(\theta)}}$ diverges while $M_\tau - N_\tau$ converge to a finite negative value.

Finally, we prove part (ii). As above, if $\lim_{\theta \rightarrow 0^+} \tau(\theta) = 0$ then the firm value is given by $\lim_{\theta \rightarrow 0^+} v(\theta) = \frac{1+\kappa^{-1}(E[v]-E[\theta])}{\rho+\lambda-b+1/\kappa}$. By monotonicity of the value function we must have $\lim_{\theta \rightarrow 0^+} v(\theta) > E[v]$, which implies $\frac{1+\kappa^{-1}(E[v]-E[\theta])}{\rho+\lambda-b+1/\kappa} > E[v]$ and $E[\theta]/\kappa < 1 - (\rho + \lambda - b)E[v]$. By using $E[v] > \underline{v}$, we obtain that a necessary condition for $\tau(0) = 0$ is that $E[\theta]/\kappa < \frac{b-a}{\rho+\lambda-a}$. Therefore, if $\kappa \leq \underline{\kappa} \equiv \frac{(\rho+\lambda-a)E[\theta]}{B}$ then $\lim_{\theta \rightarrow 0^+} \tau(\theta) > 0$. Notice that $E[\theta] < \underline{v}$ guarantees that $\bar{\kappa} > \underline{\kappa}$. \parallel

Proof. (of Proposition 6) We first prove part (i). Using the first-order condition in equation (37) we have that $\lim_{\kappa \rightarrow \infty} M(\tau; \kappa) - N(\tau, \theta; \kappa) = 1 - (\rho + \lambda - b)e^{B\tau} (\lim_{\kappa \rightarrow \infty} v(\theta; \kappa) - \theta)$. Then solving the first-order condition immediately implies that $\lim_{\kappa \rightarrow \infty} \tau(\theta; \kappa) = \tilde{\tau}(\theta)$ and $\lim_{\kappa \rightarrow \infty} v(\theta; \kappa) = \tilde{v}(\theta)$ where the functions $\tilde{\tau}(\cdot)$ and $\tilde{v}(\cdot)$ are given by equations (23)–(22).

We now prove part (ii). We first establish the following:

Lemma 3. $\forall \theta > 0$ if $0 < \lim_{\kappa \rightarrow 0} \tau(\theta; \kappa)$ then $\lim_{\kappa \rightarrow 0} \frac{\partial v(\theta; \kappa)}{\partial \theta} = \lim_{\kappa \rightarrow 0} \frac{\partial \tau(\theta; \kappa)}{\partial \theta} = 0$.

We prove the first limit in this lemma using the envelope theorem and the assumed lower bounds for the limit of τ : $\lim_{\kappa \rightarrow 0} \frac{\partial v(\theta; \kappa)}{\partial \theta} = \lim_{\kappa \rightarrow 0} 1/(1 - e^{(\rho+\lambda-b+1/\kappa)\tau(\theta; \kappa)}) = 0$. To prove the second limit in the lemma we take the limit of the first-order condition for τ , namely equation (37),

$$\frac{\partial \tau(\theta; \kappa)}{\partial \theta} = \frac{(\rho + \lambda - b + 1/\kappa)e^{(B-1/\kappa)\tau(\theta; \kappa)} \left(\frac{\partial v(\theta; \kappa)}{\partial \theta} - 1 \right)}{-B + \kappa^{-1} (1 - (\rho + \lambda - b)e^{B\tau(\theta; \kappa)} (E[v(\theta; \kappa)] - \theta))},$$

The limit of the numerator is zero given the result we just proved on the derivative of the value function. The limit of the denominator is infinite given that for each θ, κ we have $\underline{v} < v(\theta; \kappa) < \bar{v}$. This concludes the proof of the lemma.

We complete the proof of the proposition by adding a second lemma, which establishes that the hypothesis of the lemma holds.

Lemma 4. We show that for each $\theta > 0$ we have $\lim_{\kappa \rightarrow 0} \tau(\theta; \kappa) > 0$.

We prove this lemma by contradiction, supposing that $\lim_{\kappa \rightarrow 0} \tau(\theta; \kappa) = 0$. Define $p(\theta) = \lim_{\kappa \rightarrow 0} e^{-\tau(\theta; \kappa)/\kappa}$. Taking the limit on both sides of equation (25) and rearranging terms gives

$$(1 - p(\theta)) \lim_{\kappa \rightarrow 0} v(\theta, \kappa) = -p(\theta)\theta + (1 - p(\theta))E \left[\lim_{\kappa \rightarrow 0} v(\theta, \kappa) - \theta' \right]$$

where we used Lebesgue dominated convergence exchanging the expected value and the limit. The case in which $p(\theta) = 1$ yields an immediate contradiction with $\theta > 0$. Consider now the complementary case in which $p(\theta) < 1$ we get

$$\lim_{\kappa \rightarrow 0} v(\theta; \kappa) = E \left[\lim_{\kappa \rightarrow 0} v(\theta'; \kappa) - \theta' \right] - \frac{p(\theta)}{1-p(\theta)} \theta .$$

Since $\forall \kappa$ we have that $v(\theta; \kappa)$ is decreasing in θ then it is easy to show that there exist a $\bar{\theta} > 0$ such that for $\theta \in [0, \bar{\theta}]$ we have $\lim_{\kappa \rightarrow 0} v(\theta; \kappa) > E \left[\lim_{\kappa \rightarrow 0} v(\theta'; \kappa) - \theta' \right] - \frac{p(\theta)}{1-p(\theta)} \theta$ arriving to a contradiction. Hence, we have shown that $\tau(\theta; \kappa) > 0$ for all $\theta \in [0, \bar{\theta}]$. Finally using that for each κ the function $\tau(\theta, \kappa)$ is strictly decreasing in κ for $\theta \leq \bar{\theta} < E[\theta]$, we have that $\lim_{\kappa \rightarrow 0} \tau(\theta, \kappa)$ is weakly decreasing, and thus $\lim_{\kappa \rightarrow 0} \tau(\theta, \kappa) > 0$ for all $\theta > 0$. This completes the proof of the second lemma. \parallel

Proof. (of Proposition 7) Using the results of Reis's (2006) proposition 4, we have that as $E[\theta] \rightarrow 0$, $\tau(\theta) \rightarrow \bar{\tau}(\theta) = \sqrt{\frac{2\theta}{B}}$.

Let $\bar{\theta} \equiv E[\theta]$. Assuming a constant strictly positive coefficient of variation $\sqrt{v} > 0$, so that $Var(\theta) = v\bar{\theta}^2$, and using the square root approximation, we have $E[\tau] = E[\tau(\bar{\theta}) + \bar{\tau}'(\bar{\theta})(\theta - \bar{\theta})] + o(\bar{\theta}) = \bar{\tau}(\bar{\theta}) + o(\bar{\theta})$ and $Var(\tau) = (\bar{\tau}'(\bar{\theta}))^2 Var(\theta) + o(Var(\theta))$ we obtain

$$\begin{aligned} \frac{Var(\tau)}{(E[\tau])^2} &= \frac{1}{2\bar{\theta}/(b-a) + o(\bar{\theta})} \left(\left(\frac{2\bar{\theta}}{b-a} + o(\bar{\theta}) \right)^{-1/2} \frac{1}{b-a} \right)^2 Var(\theta) + \frac{o(v\bar{\theta}^2)}{2\bar{\theta}/(b-a) + o(\bar{\theta})} \\ &= \frac{1}{4} \frac{Var(\theta)}{(E[\theta])^2} + o(\bar{\theta}). \end{aligned}$$

Proof. (of Proposition 8) The derivative of the objective function in the Bellman equation (28) with respect to τ is equal to $e^{-(\rho+\lambda-a)\tau} [M(\tau) - N(\tau, \zeta)]$ where these functions are given by:

$$M(\tau) \equiv 1 - (\rho + \lambda - b)e^{B\tau} [E[v] - E[\theta]] \quad \text{and}$$

$$N(\tau, \zeta) \equiv (\rho + \lambda + \gamma - b)e^{(B-\gamma)\tau} [E[\theta] - \zeta] . \tag{38}$$

First, we note two properties of M and N : (a) if $\gamma > 0$, $\lim_{\tau \rightarrow \infty} M(\tau) - N(\tau, \zeta) = -\infty$ for any ζ ; (b) N is strictly decreasing in ζ for all $\tau < \infty$, while M does not vary with ζ . We distinguish two cases. The first case is if $M(0) - N(0, \zeta) > 0$ then (b) implies that $M(0) - N(0, \zeta) > 0$ for all $\zeta > \zeta^*$. The first-order condition then implies that $\tau(\zeta) > 0$ for all ζ and $\zeta^* = \zeta$. If instead, $M(0) - N(0, \zeta) \leq 0$, then $\zeta^* \geq \zeta$ with $\tau(\zeta) = 0$ and $v(\zeta) = E[v] - \zeta$ for all $\zeta < \zeta^*$. For $\zeta \geq \zeta^*$, we have that $M(0) - N(0, \zeta) \geq 0$ which when combined with (a) it implies that the solution to the firm problem is given by the solution to the first-order condition at a value $0 < \tau(\zeta) < \infty$. Moreover, by differentiating the first-order condition at the interior maximum we obtain

$$\frac{\partial \tau(\zeta)}{\partial \zeta} = \frac{N_{\zeta}}{M_{\tau} - N_{\tau}} > 0,$$

where $N_{\zeta}(\tau; \zeta) = -(\rho + \lambda + \gamma - b)e^{-(\gamma-b+a)\tau} < 0$, and $M_{\tau} - N_{\tau} \leq 0$ by the definition of a maximum. The second case is if $M(0) - N(0, \zeta) > 0$. In this case, the optimal $\tau(\zeta)$ is always strictly bigger than zero. Differentiating the first-order condition with respect to τ we have $S(\tau) \equiv M_{\tau}(\tau) - N_{\tau}(\tau, \zeta) = -B + \gamma (1 - (\rho + \lambda - b)e^{B\tau} [E[v] - E[\theta]])$. We prove now that the objective function has at most one interior local maximum. Let us denote by τ_1 the smallest local maximum. Notice that $S(\tau_1) \leq 0$ by definition of a local maximum. If there would be another local maximum $\tau_2 > \tau_1$, then there must be a value of $\tau_m \in (\tau_1, \tau_2)$ that is a local minimum requiring $S(\tau_m) \geq 0$. But notice that the function $S(\tau)$ is strictly decreasing in τ which is a contradiction of $S(\tau_m) \geq 0 > S(\tau_1)$ as $\tau_m > \tau_1$.

To prove point that $\tau(\zeta)$ is not bounded assume for a contradiction that $\tau(\zeta)$ has an upper bound. Then, provided that $\gamma < \infty$, there exists a ζ large enough for which the value function is arbitrarily negative, and in particular smaller than \underline{v} , implying that the upper bound is not optimal. Once we have established that $\lim_{\zeta \rightarrow \infty} \tau(\zeta) = \infty$.

To prove point (ii) assume by contradiction that $\underline{\zeta} \geq 0$ and $\tau(\underline{\zeta}) = 0$ then: if $\underline{\zeta} > 0$, $v(\underline{\zeta}) = E[v] - \underline{\zeta} < E[v]$, which is a contradiction because $v(\underline{\zeta}) \geq E[v]$ as $v(\underline{\zeta})$ is decreasing in ζ ; if $\underline{\zeta} = 0$, then $v(\underline{\zeta}) = E[v]$ which is only possible if $v(\zeta)$ is constant. But this is possible only if $\tau(\zeta) = \infty$ for all ζ , which is contradiction given the arguments above. By definition of ζ^* , given that $\tau(\zeta) > 0$ it follows that $\zeta^* = \underline{\zeta}$.

Finally notice that a stochastically higher distribution of cost \hat{G} implies a weakly lower value function $v(\zeta)$ for each ζ and hence a lower value of $E[v]$. For any interior optimum we have

$$\frac{\partial \tau(\zeta)}{\partial E[v]} = - \frac{M_{E[v]}}{M_{\tau} - N_{\tau}} < 0,$$

since $M_{\tau} - N_{\tau} < 0$ and $M_{E[v]} < 0$. \parallel

Proof. (of Proposition 9) We first prove the part (i) of the proposition in steps.

- (1) As $\gamma \rightarrow 0$ in the first-order condition in equation (38) we have that

$$\lim_{\gamma \rightarrow 0} M(\tau; \gamma) - N(\tau, \zeta; \gamma) = e^{-(\rho+\lambda-a)\tau} (1 - e^{B\tau} [\rho + \lambda - b][E[v] - \zeta]). \quad (39)$$

equation (29) is then obtained by setting the expression above equal to zero for a finite positive value of τ . If this expression is negative at $\tau = 0$, then we set $\tau = 0$, while we set $\tau = \infty$ if it is strictly positive for all finite τ .

- (2) Using the optimal policy in equation (29) in the Bellman equation (28) the value function takes the following form

$$v(\zeta) = \begin{cases} E[v] - \zeta & \text{if } \zeta \leq \zeta \leq E[v] - \bar{v} \\ \underline{v} + B\underline{v}\bar{v} \left(\frac{E[v] - \zeta}{\bar{v}} \right)^{\frac{\rho+\lambda-a}{B}} & \text{if } E[v] - \bar{v} < \zeta < E[v] \\ \underline{v} & \text{if } \zeta \geq E[v] \end{cases} \quad (40)$$

Integrating both sides of equation (40) with respect to \hat{G} we obtain

$$E[v] = \underline{v} \left(1 - \hat{G}(E[v] - \bar{v}) \right) + B\underline{v}\bar{v} \int_{E[v] - \bar{v}}^{E[v]} \left(\frac{E[v] - x}{\bar{v}} \right)^{\frac{\rho+\lambda-a}{B}} d\hat{G}(x) + \bar{v} \int_{\zeta}^{E[v] - \bar{v}} \left(\frac{E[v] - x}{\bar{v}} \right) d\hat{G}(x). \quad (41)$$

- (3) If $\zeta \geq 0$ then $\hat{\tau}(\zeta) > 0$. Note that $E[v] < \bar{v}$ since the right hand side is the value of the case of not observation cost. Hence $\hat{\tau}(\zeta) > 0$, as we argued for the general case.
 (4) If $\zeta \geq 0$ equation (41) becomes

$$E[v] = \underline{v} + B\underline{v}\bar{v} \int_{\zeta}^{E[v]} \left(\frac{E[v] - \zeta}{\bar{v}} \right)^{\frac{\rho+\lambda-a}{B}} d\hat{G}(\zeta). \quad (42)$$

The first and second derivatives of the right-handside of the last equation w.r.t $E[v]$ are

$$1 > [(\rho + \lambda) - b] \frac{\rho + \lambda - a}{B} - 1 \int_{\zeta}^{E[v]} (E[v] - \zeta)^{\frac{\rho + \lambda - a}{B} - 1} d\hat{G}(\zeta) \geq 0, \\ [(\rho + \lambda) - b] \frac{\rho + \lambda - a}{B} - 1 \left(\frac{\rho + \lambda - a}{B} - 1 \right) \int_{\zeta}^{E[v]} (E[v] - \zeta)^{\frac{\rho + \lambda - a}{B} - 2} d\hat{G}(\zeta) \geq 0.$$

Note that the first derivative is zero at $E[v] = \zeta$. If $E[v] < 1/(\rho + \lambda - b)$ the first derivative is strictly increasing and strictly smaller than one. Hence, there is at most one solution for equation (42) with $E[v] < 1/(\rho + \lambda - b)$. If there is a second intersection where $E[v]$ satisfies equation (42) it must be at a point where $E[v] > 1/(\rho + \lambda - b)$ since the slope must be larger than one. Given that the right-hand side of equation (42) is convex in $E[v]$ there are at most two intersection. Now we show that there is at least one intersection. Consider the case where the distribution \hat{G} is concentrated at $\zeta = 0$. Then there is a unique intersection at $E[v] = \bar{v}$. Next, for any other \hat{G} , fixing a given $E[v]$, the right-hand side of equation (42) is smaller than in the case of \hat{G} concentrated. Hence, there is always an intersection with slope smaller than one.

- (5) The solution of the foc is a maximum. Consider a case where $E[v] - \zeta < \zeta < E[v]$, then the second derivative of the right-hand side of the Bellman equation, evaluated at equation (29) gives:

$$e^{-(\rho+\lambda-a)\tau(\zeta)} \left(-(\rho + \lambda - a) + e^{B\tau(\zeta)} [-(\rho + \lambda) + b]^2 [E[v] - \zeta] \right) = e^{-(\rho+\lambda-a)\tau(\zeta)} B < 0.$$

- (6) Since $\mathbb{E} \left[\max(E[v] - \zeta, 0)^{\frac{\rho+\lambda-a}{B}} \right]$ is the expected value of a convex function of ζ for a fixed value of $E[v]$, a mean preserving spread in $\hat{G}(\cdot)$ increases its value. Hence, a mean preserving spread in the distribution of ζ increases the rhs of equation (42) for each $E[v]$, and thus increases the value of the intersection. \parallel

Proof. (of Proposition 10) Let $\bar{\theta} \equiv \mathbb{E}(\theta)$ and $\bar{\tau} = \tau(\bar{\theta})$. Integrating the value function in equation (28) for the case of zero variance of observation cost gives

$$v(\bar{\theta}) - \bar{\theta} = \frac{(1 - e^{-(\rho+\lambda-a)\bar{\tau}})/(\rho + \lambda - a) - \bar{\theta}}{1 - e^{-(\rho+\lambda-b)\bar{\tau}}}$$

Combining the right-hand side of this equation with equation (29) allows us to write $v(\bar{\theta}) - \bar{\theta} = \frac{e^{-(b-a)\bar{\tau}}}{\rho + \lambda - b}$ which then gives

$$\tau'(\bar{\theta}) = \frac{\rho + \lambda - b}{b - a} e^{(b-a)\bar{\tau}} = \frac{\rho + \lambda - b}{b - a} e^{\sqrt{2\bar{\theta}(b-a)}} = \frac{\rho + \lambda - b}{\eta(\eta - 1)\sigma^2/2} e^{\sqrt{2\bar{\theta}\eta(\eta-1)\sigma^2/2}} \quad (43)$$

where the last equality uses the square root formula derived above for the limit case of no variance. Next, assume a constant strictly positive coefficient of variation $\sqrt{v} > 0$, so that $\text{Var}(\theta) = v\bar{\theta}^2$. When $v \approx 0$, $\tau(\theta)$ is approximately given by $\sqrt{2\bar{\theta}/B}$ so that $(\mathbb{E}(\tau))^2 = 2\bar{\theta}/(b-a) + o(\bar{\theta})$. Using equation (43), and $\text{Var}(\tau) = (\tau'(\bar{\theta}))^2 \text{Var}(\theta) + o(\text{Var}(\theta))$ we obtain

$$\begin{aligned} \frac{\text{Var}(\tau)}{(\mathbb{E}(\tau))^2} &= \left(\frac{\rho + \lambda - b}{b-a} \right)^2 \left(1 + \sqrt{2\bar{\theta}(b-a) + o(\bar{\theta})} + \frac{1}{2!} (2\bar{\theta}(b-a) + o(\bar{\theta})) + \dots \right)^2 \frac{\text{Var}(\theta)}{2\bar{\theta}/(b-a) + o(\bar{\theta})} \\ &\quad + \frac{o(v\bar{\theta}^2)}{2\bar{\theta}/(b-a) + o(\bar{\theta})} \\ &= \left(\frac{\rho + \lambda - b}{(b-a)} \right)^2 \left(1 + o(\sqrt{\bar{\theta}}) \right) \frac{\text{Var}(\theta)}{2\bar{\theta}/(b-a) \left(1 + \frac{o(\bar{\theta})}{2\bar{\theta}/(b-a)} \right)} + o(\bar{\theta}) \\ &= \left(\frac{\rho + \lambda - b}{(b-a)} \right)^2 \frac{\text{Var}(\theta)}{\bar{\theta}^2} \frac{(b-a)\bar{\theta}}{2} \left(1 + o(\sqrt{\bar{\theta}}) \right) \left(1 - \frac{o(\bar{\theta})}{\bar{\theta}} \right) + o(\bar{\theta}) \\ &= (\rho + \lambda - b)^2 v \frac{\bar{\theta}}{2(b-a)} \left(1 + o(\sqrt{\bar{\theta}}) - \frac{o(\bar{\theta})}{\bar{\theta}} - \frac{o(\bar{\theta}^{\frac{3}{2}})}{\bar{\theta}} \right) + o(\bar{\theta}) \\ &= (\rho + \lambda - b)^2 v \frac{\bar{\theta}}{2(b-a)} + o(\bar{\theta}) \end{aligned}$$

which, using $v = \frac{\text{Var}(\theta)}{(\bar{\theta})^2} > 0$, gives equation (30).

Proof. (of Proposition 11) The cumulative distribution function of observation cost is given by $\hat{G}(\zeta; E[v]) = 1 - K(\tau(\zeta; E[v]))$. By differentiating the last equation with respect to ζ , we obtain

$$\hat{g}(\zeta; E[v]) = -K'(\tau(\zeta; E[v])) \frac{\partial \tau(\zeta; E[v])}{\partial \zeta} \quad \text{for all } \zeta > \underline{\zeta}. \quad (44)$$

Then by replacing the optimal policy and its derivative we obtain

$$\begin{aligned} \hat{g}(\zeta; E[v]) &= \frac{\xi \exp(-\xi \tau(\zeta))}{B(E[v] - \zeta)} \\ &= \frac{\xi \exp\left(\frac{\xi}{B} \log([\rho + \lambda - b][E[v] - \zeta])\right)}{B[E[v] - \zeta]} \\ &= a_0 \frac{\exp(\log(a_1^{a_0} [E[v] - \zeta]^{a_0}))}{[E[v] - \zeta]} = a_0 \frac{a_1^{a_0} [E[v] - \zeta]^{a_0}}{[E[v] - \zeta]} = a_0 a_1^{a_0} [E[v] - \zeta]^{a_0 - 1} \end{aligned}$$

where $a_0 = \frac{\xi}{B}$ and $a_1 = \rho + \lambda - b$.

$$E[v] = \underline{v} + B \underline{v} \bar{v} \int_{E[v] - \bar{v}}^{E[v]} \left(\frac{E[v] - \zeta}{\bar{v}} \right)^{\frac{\rho + \lambda - a}{B}} \hat{g}(\zeta; E[v]) d\zeta. \quad (45)$$

Finally, by using equation (31) we obtain

$$E[v] = \underline{v} + B \underline{v} \bar{v} \frac{\xi}{B} \int_{E[v] - \bar{v}}^{E[v]} \left(\frac{E[v] - \zeta}{\bar{v}} \right)^{\frac{\rho + \lambda - a + \xi}{B} - 1} d\zeta, \quad (46)$$

which gives

$$E[v] = \underline{v} + B \underline{v} \bar{v} \frac{\xi \underline{v}}{1 + \xi \underline{v}} = \underline{v} + (\bar{v} - \underline{v}) \frac{\xi \underline{v}}{1 + \xi \underline{v}}.$$

The distribution ζ is then a displaced Beta distribution, *i.e.* $\zeta = c + dz$ where z has a standard beta distribution with shape parameters $(\tilde{a}, \tilde{b}) = 1, \xi/B$. The displaced parameters are $c = E[v] - \bar{v}$ and $d = \bar{v}$. Hence the coefficient of variation of ζ , denoted by $cv(\zeta)$, is:

$$cv(\zeta) = \sqrt{\frac{\xi/B}{2 + \xi/B} \frac{1}{1 - (1 + \xi/B)(\bar{v} - E[v])/\bar{v}}},$$

and using the expression for $E[v]$ we obtain $\bar{v} - E[v] = (\bar{v} - \underline{v}) / (1 + \xi \underline{v})$, and then

$$cv(\zeta) = \sqrt{\frac{\xi/B}{2 + \xi/B}} \frac{1}{1 - \left(\frac{1 + \xi/B}{1 + \xi \underline{v}}\right) \left(\frac{\bar{v} - \underline{v}}{\underline{v}}\right)},$$

which using that $\underline{v} = (1/B)(\bar{v} - \underline{v})/\bar{v}$ then becomes

$$cv(\zeta) = \sqrt{\frac{\xi/B}{2 + \xi/B}} \left(\frac{1 + (\xi/B)(\bar{v} - \underline{v})/\bar{v}}{\underline{v}/\bar{v}}\right) = \sqrt{\frac{\xi/B}{2 + \xi/B}} \left(\frac{1 + \xi \underline{v}}{\underline{v}/\bar{v}}\right)$$

||

Acknowledgments. We thank Carlos Carvalho, Xavier Gabaix, Christian Hellwig, Pat Kehoe, Herve Le Bihan, Bartosz Mackowiak, Ricardo Reis, Victor Rios-Rull as well as seminar participants at UCL, EIEF, the Bank of England, ECB, International Network on Expectations and Coordination at NYU, NBER Summer Institute 2012, "Rational Inattention and Related Theories" Prague 2012 and Oxford 2014, SED Annual meeting 2015, Tor Vergata University, and T2M Conference at Lausanne University for their comments. Part of the research for this paper was sponsored by the ERC advanced grant 324008. Philip Barrett and Jean Fleming provided excellent research assistance.

Supplementary Data

Supplementary data are available at *Review of Economic Studies* online.

REFERENCES

- ABEL, A. B., EBERLY, J. C. and PANAGEAS, S. (2007), "Optimal Inattention to the Stock Market", *American Economic Review*, **97**, 244–249.
- ALVAREZ, F. and LIPPI, F. (2014), "Price Setting with Menu Cost for Multiproduct Firms", *Econometrica*, **82**, 89–135.
- ALVAREZ, F. E., BOROVIKOVA, K. and SHIMER, R. (2015), "A Nonparametric Variance Decomposition Using Panel Data" (Mimeo, University of Chicago).
- ALVAREZ, F. E., LE BIHAN, H. and LIPPI, F. (2014), "Small and Large Price Changes and the Propagation of Monetary Shocks", (Working Paper 20155, National Bureau of Economic Research, NBER).
- ALVAREZ, F. E., LIPPI, F. and PACIELLO, L. (2011), "Optimal Price Setting with Observation and Menu Costs", *Quarterly Journal of Economics*, **126**, 1909–1960.
- ALVAREZ, F. E., LIPPI, F. and PACIELLO, L. 2015. "Phillips Curves with Observation and Menu Costs", (Working Paper No. 8/2015, EIEF).
- BONOMO, M. and CARVALHO, C. (2004). "Endogenous Time-Dependent Rules and Inflation Inertia", *Journal of Money, Credit and Banking*, **36**, 1015–1041.
- CABALLERO, R. J. (1989), "Time Dependent Rules, Aggregate Stickiness And Information Externalities", (Discussion Papers 198911, Columbia University).
- CARVALHO, C. and SCHWARTZMAN, F. (2012), "Selection and Monetary Non-Neutrality in Time-Dependent Pricing Models", (Federal Reserve Bank of Richmond, Technical Report).
- GOLOSOV, M. and LUCAS, Jr., R. E. 2007. "Menu Costs and Phillips Curves." *Journal of Political Economy*, **115**, 171–199.
- KARLIN, S. and TAYLOR, H. M. (1998), *An Introduction to Stochastic modelling*, Vol. 1 (Academic Press, Elsevier).
- MACKOWIAK, B. and WIEDERHOLT, M. (2009), "Optimal Sticky Prices under Rational Inattention", *American Economic Review*, **99**, 769–803.
- MANKIW, N. G. and REIS, R. (2002), "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve", *The Quarterly Journal of Economics* **117**, 1295–1328.
- NAKAMURA, E. and STEINSSON, J. (2008), "Five Facts about Prices: A Reevaluation of Menu Cost Models", *The Quarterly Journal of Economics*, **123**, 1415–1464.
- REIS, R. (2006), "Inattentive Producers", *Review of Economic Studies*, **73**, 793–821.
- SIMS, C. A. (2003), "Implications of Rational Inattention", *Journal of Monetary Economics*, **50**, 665–690.
- WOODFORD, M. (2009), "Information-Constrained State-Dependent Pricing", *Journal of Monetary Economics*, **56**, S100–S124.
- ZBARACKI, M. J., RITSON, M., LEVY, D., *et al.* (2004), "Managerial and Customer Costs of Price Adjustment: Direct Evidence from Industrial Markets", *The Review of Economics and Statistics* **86**, 514–533.