Working Paper Series n° 597

March, 2010
Revised January, 2011

"A "quantized" approach to rational inattention"

Gilles SAINT-PAUL

Institut d'Economie Industrielle (IDEI) – Manufacture des Tabacs
Aile Jean-Jacques Laffont – 21, allée de Brienne – 31000 TOULOUSE – FRANCE
Tél. + 33(0)5 61 12 85 89 – Fax + 33(0)5 61 12 86 37 – **www.idei.fr** – contact@cict.fr

INSTITUT D'ECONOMIE INDUSTRIELLE

A "quantized" approach to rational inattention

Gilles Saint-Paul*
Toulouse School of Economics

January 10, 2011

ABSTRACT

In this paper, I propose a model of rational inattention where the choice variable is a deterministic function of the exogenous variables, and still only a finite amount of information is being used. This holds provided the choice variable is discrete rather than continuous; that is, the mapping from the realization of the exogenous variables to the endogenous ones is piece-wise constant.

Thus, limited information is now a source of lumpiness in behavior, rather than a source of noise. A central result is that the mutual information between the exogenous variable and the endogenous one is simply equal to the entropy, in the usual discrete sense, of the endogenous variable.

The approach is illustrated with two applications: a general linear-quadratic problem with a uniform distribution, and a simple static model of price-setting where individual price setters face aggregate monetary shocks and idiosyncratic productivity shocks.

Keywords: Rational inattention, lumpy adjustment, price-setting, monetary policy, mutual information, entropy

JEL: D8, E3

*I am indebted to Enduardo Engel, Giuseppe Moscarini, Filip Mateka, Christopher Sims, and seminar participants at Sciences Po, Paris, CREI, Universitat Pompeu Frabra, Barcelona, Universidade do Minho, Braga, Tilburg, and IMT Lucca, for helpful comments and suggestions.

1 Introduction

In a series of seminal papers, Sims (2003,2006) has proposed a novel approach to bounded rationality. It is based on the view that people face information capacity constraints defined using Shannon's (1948) theory of information. More precisely, the number of bits that one can use to process the exogenous variables (like income) into the endogenous ones (like consumption) is limited. That informational requirement is defined by Shannon's mutual information concept, which tells us the amount of information obtained on a variable when one observes another, correlated one.

As a consequence of that information constraint, the endogenous variable is noisy compared to the optimal behaviour that would prevail absent an informational constraint. Thus, is many applications (such as Luo (2008), who studies a consumption problem, and the papers cited below on price-setting) the agents rationally allocates this noise so as to maximize its utility subject to the information capacity constraint. The more noisy is the endogenous variable in a given zone of the distribution of exogenous variables, the less the agent pays attention to that zone and the greater the informational capacity left for processing other zones.

As pointed out by Sims, the reason why noise must inevitably arise is that if the distribution of the exogenous variables is continuous, then an infinite amount of information would be needed to process a deterministic mapping from the exogenous variables into the endogenous ones.

In some settings, the noise is inherent to the problem of measuring a signal and the agents' informational capacity is used to reduce such a noise. The rational inattention theory then tells us, in some sense, how to optimally design the noise so as to get the highest possible welfare subject to the information capacity constraint.

In other settings, though, the result that behaviour adds noise to the exogenous variables is unpalatable. If the realization of the latter is perfectly observed then the agent would have to generate the noise artificially, but then it is problematic to ignore the information needed to generate such noise. It would then be more reasonable to assume that the behaviour of the agent remains deterministic while the information processing constraint prevents it from targeting the optimal behaviour. However, that is not what is happening in the rational inattention literature.

In this paper, I propose an alternative approach to that issue. The idea is that the choice variable may be a deterministic function of the exogenous one and still make use of a finite amount of information if the choice variable is discrete rather than continuous; that is, the mapping from the realization of the exogenous variables to the endogenous ones is piece-wise constant, reflecting the fact that the agent can only elect a finite number of values for the choice variable, because of the informational constraint.

Thus, limited information is now a source of lumpiness in behavior, rather than a source of noise. The state space faced by the agent is partitioned into clusters and all points in the same cluster yield the same action. Of course, limited information is not the only source of lumpy behavior; it is well known that there are other sources, such as fixed or linear adjustment costs. But the approach proposed here yields many potentially testable predictions: In general, we expect that the greater the information processing ability of an economic entity, the less lumpy its behavior.

Another central result (Section 2) is that the mutual information between the exogenous variable and the endogenous one is simply equal to the entropy, in the usual discrete sense, of the endogenous variable. That is, the mutual information does not depend on the exact mapping from the exogenous variable to the endogenous one but only on the probability weights of the (discrete) distribution of the latter. This remedies some weaknesses of the continuous notions of entropy which is used in the literature, which makes it impossible to separate the, probability weights of the variables from their actual values.

Sections 3 and 4 illustrate the kind of results that my approach would deliver by applying it to two simple examples: a general linear-quadratic problem with a uniform distribution, and a simple static model of price-setting where individual price setters face aggregate monetary shocks and idiosyncratic productivity shocks. This model delivers a lumpy price-setting behavior where the number and size of clusters depends on the dispersion of shocks and the firm's information processing capacity. It is consistent with recent micro-level evidence on reference prices found by Eichenbaum et al (2008).

The literature that has studied this issue (in particular, Mackowiak and Wiederholt (2009a,b), Paciello (2007)) uses Sim's noisy approach and has shown that under rational inattention prices were "sticky" in the sense that the aggregate price level was not reacting one for one to the aggregate money stock[1]. However, prices are not lumpy: even a small monetary shock will generate a small (but non neutral) response of individual prices (one notable exception is Moscarini (2004))[2]. Here, in contrast, rational inattention leads to lumpy price-setting behavior; for prices to change, the shocks faced by a firm must be large enough to trigger a move to a different cluster. As discussed in Section 5, this makes a substantial difference. In the noisy approach, it is optimal to react less than one for one to monetary shocks because one can only react to a noisy measure of those shocks, as in the Lucas (1972) misperception model. Consequently, such underreaction also prevails at the aggregate level. Here, though, no noise is introduced and stickiness arises at the individual level because the same price is charge within a cluster of realizations of the individual price setter's relevant shock. But, as in Caplin and Spulber (1986), such stickiness is greatly reduced in the aggregate because of the contribution of the firms which move across clusters

[1] The same result is reached by Saint-Paul (2005) in a world where firms are irrational and experiment alternative price-setting rules, while exerting local spillovers on each other.

[2] In that paper, lumpiness arises for different reasons than here. Time is continuous and there is a constraint on the flow of information processed by the agent. the exogenous variable follows a diffusion process. A noisy signal of that variable can be obtained at a cost. The cost structure of information is such that the signal will be drawn infrequently, at discrete dates. Thus there is lumpiness "in time" rather than in the state space. Consequently, the model is similar to that of Mankiw and Reis's (2002) sticky information paper.

as a result of a monetary shock.

It is important to note that while lumpiness is the only way to reconcile limited mutual information with deterministic behavior, the converse is not true. Depending on the objective function and the distribution of noise, the support of the endogenous variable may either be discrete or continuous, as recently found by Matejka (2008) and Matejka and Sims (2010), who provide some partial but powerful sufficient conditions for discreteness to arise. Yet, even in this case, it is always optimal to have a noisy behaviour, so that the implications for aggregate price stickiness will resemble those of Mackowiak and Wiederholt (2009a).

Section 6 provides a more general discussion and concludes.

2 Continous and discrete entropy and mutual information

It is somewhat important to realize that there are two different concepts of entropy. For a discrete distribution with n outcomes and probabilities p_i, $i = 1, ..., n$, we may define entropy as

$$S = -\sum_{i=1}^{n} p_i \log p_i.$$

On the other hand, for a continuous distribution with density $f(x)$, we may define entropy as

$$H(f) = -\int f(x) \log f(x).dx.$$

The reason why the two concepts do not coincide is as follows. A discrete distribution is always the limit of a sequence of continuous distributions, as they become more concentrated around the discrete outcomes. However, the continuous entropy H of those approximations does not converge to the corresponding S. Instead, it converges to $-\infty$.

Take for example the extreme case where $x = 0$ with probability 1. Clearly, $S(x) = 0$. This discrete distribution is the limit of the continuous one defined by density $f_\varepsilon(x) = f(x/\varepsilon)/\varepsilon$, for any density $f()$ over $(-\infty, +\infty)$, which is regular and has $f(0) > 0$, as ε goes to zero (in terms of distribution theory, these distributions converge to a Dirac function $\delta(x)$). Furthermore,

$$H(f_\varepsilon) = H(f) + \log \varepsilon,$$

so that

$$\lim_{\varepsilon \to 0} H(f_\varepsilon) = -\infty.$$

Entropy is lower, the more concentrated the distribution. For both discrete and continuous distributions, the most concentrated one is when all the mass is at a single point. But the lower bound of S is zero, while that of H is -∞.

Let us now turn to mutual information, which plays a key role in the theory of rational inattention. We consider two random variables x and y. Their densities are $g()$ and $f()$, respectively. For any realization of x, we denote by $f(y \mid x)$ the conditional distribution of y and its entropy is

$$H(y \mid x) = -\int_y f(y \mid x) \log f(y \mid x).dy.$$

This can be averaged over x, which allows to define the conditional entropy of y :

$$H_x(y) = \int_x H(y \mid x) g(x) dx.$$

Now it can be easily shown that the entropy of the joint distribution of x and y, $H(x, y)$, is such that[3]

$$H(x, y) = H(x) + H_x(y) = H(y) + H_y(x).$$

Consequently, we have that

$$H(y) - H_x(y) = H(x) - H_y(x) = M(x, y),$$

[3] In fact, that property is one of the axioms imposed by Shannon to derive his functional form for entropy.

which is the *mutual information* between x and y. This quantity tells us how much knowledge of one variable reduces the entropy of the other, on average. If the two variables are independent, then $M(x, y) = 0$. On the other hand, if one had $y = x$, then the joint distribution is degenerate and $f(y \mid x)$ becomes equal to the Dirac function $\delta(y - x)$. Hence all the $H(y \mid x)$ are equal to $-\infty$ and so is $H_x(y)$. We then have that $M(x, y) = +\infty$. This means that knowledge of x gives us an infinite amount of information about y. The same conclusion would be reached if instead of $y = x$, there was any other mapping which allowed to retreive one variable from the other.

The theory of rational inattention, as proposed by Sims, assumes that an agent receives a signal y (say, income), which must be processed into a decision variable x (say, consumption). The agent's ability to process information is limited and that limit takes the form of a constraint on the mutual information between the two variables:

$$M(x, y) \leq K.$$

Since $M(x, y) = +\infty$ if x is a deterministic function of y, this constraint cannot be matched. The endogenous variable must be related to the exogenous one in a noisy fashion for the information capacity constraint to be matched. In other words, processing a continuum of real values with perfect precision requires an infinite amount of information.

I now show that there is an important exception to that principle, and this is the case when x, while being a deterministic function of y, only takes a finite number values. There is then no longer a mapping from y to x. While x can be retreived from y, the converse is not true. In such a case, the mutual information between x and y remains finite, and is in fact equal to the discrete entropy S of the random variable x.

Let us consider a collection of values of x, $X = \{x_1, ..., x_n\}$, and assume that any y is assigned to one of those values, called $x(y)$. For any $x \in X$, we define $T_x = \{y, x(y) = x\}$. To avoid manipulating infinite quantities, I will consider my deterministic assignment as the limit, for $\varepsilon \to 0$, of the random variable x defined by its conditional distribution:

$$f_\varepsilon(x \mid y) = \frac{1}{\varepsilon}\hat{f}\left(\frac{x - x(y)}{\varepsilon}\right). \tag{1}$$

Here again, $\hat{f}()$ is any regular density such that $\hat{f}(0) > 0$. We are again in a situation where the conditional of x is a Dirac, now around $x(y)$, and we approximmate it by a density which becomes increasingly concentrated around $x(y)$. To fix ideas, one can just take the standard normal density for $\hat{f}()$.

Then the following can be proved:

Theorem 1 – Let $M(\varepsilon)$ be the mutual information between x and y if x is distributed according to (1). Let $S(X) = -\sum_i P_i \log P_i$ be the discrete entropy of the random variable whose realization is x_i, with corresponding probability $P_i = F(T_i)$. Then

$$\lim_{\varepsilon \to 0} M(\varepsilon) = S(X).$$

Proof – See Appendix.

The conditional entropy of x, if x is distributed as f_ε, is

$$\begin{aligned} H_y(x;\varepsilon) &= -\int_y f(y)\int_x \frac{1}{\varepsilon}\hat{f}\left(\frac{x - x(y)}{\varepsilon}\right)\log\left[\frac{1}{\varepsilon}\hat{f}\left(\frac{x - x(y)}{\varepsilon}\right)\right]dxdy \\ &= -\int_y f(y)\int_z \hat{f}(z)\log\left[\frac{1}{\varepsilon}\hat{f}(z)\right]dzdy \\ &= H(\hat{f}) + \log\varepsilon. \end{aligned}$$

So, clearly, $\lim_{\varepsilon\to 0} H_y(x;\varepsilon) = -\infty$.

Consider now the entropy of x, given ε, denoted by $H(x;\varepsilon)$. The unconditional density of x is

$$g_\varepsilon(x) = \sum_{i=1}^{n} \frac{1}{\varepsilon}\hat{f}(\frac{x - x_i}{\varepsilon})F(T_i).$$

Thus,

$$\begin{aligned}
H(x;\varepsilon) &= -\int_x g_\varepsilon(x)\log g_\varepsilon(x).dx \\
&= -\sum_{i=1}^{n}\int_x \frac{1}{\varepsilon}\hat{f}(\frac{x-x_i}{\varepsilon})F(T_i)\log\left(\frac{1}{\varepsilon}\hat{f}(\frac{x-x_i}{\varepsilon})F(T_i)\right)dx \\
&= -\sum_{i=1}^{n}\int_z \hat{f}(z)F(T_i)\log\left(\frac{1}{\varepsilon}\hat{f}(z)F(T_i)\right)dz \\
&= H(\hat{f})+\log(\varepsilon)+S(X),
\end{aligned}$$

where $S(X)$ is the discrete entropy of the random variable whose realization is x_i, with corresponding probability $p_i = F(T_i)$. Thus $H(x;\varepsilon)$ also converges to $-\infty$ when ε becomes nil, however, the mutual information remains finite and

$$M(x,y;\varepsilon) = S(X).$$

This is independent of ε and equal to the discrete entropy of the random variable x_i. Obviously, it remains equal to that as $\varepsilon \to 0$. Hence the mutual information of our assignment process is finite.

3 The linear-quadratic case

I now apply these ideas to the linear-quadratic case. In its simplest case, the agent receives a continuous signal y with density $f(y)$ and associated measure $F(M) = \int_M f(y)dy$, and wants to approximate it (in the least squares sense) by a deterministic function $x(y)$ which is constant over each subset of a finite partition of the domain of y. Thus, using the preceding derivation for mutual information in the discrete case, we can formulate the problem as follows (in the sequel I will use natural logarithms in the definition of entropy. Thus K is expressed in bits / $\ln 2$).

$$(\mathrm{P}) \quad : \quad \min_{n,S=(x_1,\ldots,x_n),x():\mathbb{R}\to S} E(x(y)-y)^2$$

$$s.t. -\sum_{i=1}^{n} F(x^{-1}(x_n))\ln F(x^{-1}(x_n)) \leq K$$

A first property, which is unsurprising given the convexity of the loss function, is that the optimum must be such that the sets $x^{-1}(x_n)$, denoted by S_n, have a convex interior.

Lemma 1 – *Consider a solution to (P). Then* $C(\mathring{S}_n) = \mathring{S}_n \cup \partial \mathring{S}_n$ *is convex.*

Proof – see Appendix.

Lemma 1 tells us that the S_n must be intervals, except for closed subsets of measure zero. In practice those subsets are irrelevant so we will focus on solution that are piece-wise constant.

Lemma 2 – $x_n = E(y \mid y \in S_n)$. *Therefore* $x(y)$ *is non decreasing.*

Proof: the first part is straightforward from the optimal choice of x_n. The second derives from the fact that the S_ns are intervals except for subsets of measure zero.

I now focus on the case where $f()$ is uniform over $[0, 1]$. It is then possible to fully characterize the equilibrium:

Theorem 2 – *Assume* $f()$ *is uniform. Then an optimal policy is such that*

(i) The interval $[0, 1]$ *is partitioned into* N *adjacent intervals* $[y_n, y_{n+1}], y_0 = 0, y_N = 1$.

(ii) $N = INT^+(e^K)$, *where* $INT^+(z)$ *is the smallest integer* m *such that* $z \leq m$.

(iii) $N - 1$ *intervals have the same length* Δ, *where* Δ *is the smallest solution to*

$$-(N-1)\Delta \ln \Delta - (1-(N-1)\Delta)\ln(1-(N-1)\Delta) = K,$$

while the remaining interval has length $1 - (N-1)\Delta$.

(iv)

$$\Delta < 1/N < 1 - (N-1)\Delta$$

(v) For $y \in [y_n, y_{n+1}], x(y) = x_n = \frac{y_n + y_{n+1}}{2}$

(v) The resulting value function is $V = (N-1)\Delta^3 + (1-(N-1)\Delta)^3$

(vi) The arrangement of those intervals is irrelevant.

Proof — See Appendix.

Note that if capacity K is such that there is an integer number of bits, then $K = k \ln 2$ with k integer, and $e^K = 2^k$. In this important special case, the optimal solution, quite naturally, consists in splitting the interval into 2^k intervals, since one needs exactly k bits to encode the actual interval to which y is assigned. Furthermore, in this limit case where the capacity constraint is marginally binding for $N = 2^k$, all intervals will have the same length $1/2^k$.

If $K/\ln 2$ is not integer, then partitioning into equal intervals is not optimal. Instead, we have one more interval than the largest number of intervals that would allow us to have an equal partition while meeting the informational constraint. We pick $N-1$ equally sized intervals of length Δ, and the remaining one has length $\Delta' = 1-(N-1)\Delta > \Delta$. Δ is such that the informational constraint binds with equality.

4 An application to price-setting

We now discuss the implications of the approach derived above for the problem of price setting and the effects of monetary policy.

Let us consider the following static version of the standard new Keynesian model[4]. There is a continuum of consumers-yeoman farmers of total mass 1. They are indexed by i and they monopolistically supply an atomistic good with the same index i. Thus there is also a continuum of goods of mass 1. The utility function for individual j is

$$V_j = E \ln \left[\left(\int_0^1 c_{ij}^{\alpha} di \right)^{\frac{1}{2\alpha}} \left(\frac{m_j}{p} \right)^{1/2} X^{-\psi} - z_j x_j^{1+\mu} \right],$$

where E is the expectations operator, c_{ij} consumption of good i, m_j money holdings, p the aggregate price level, z_j an idiosyncratic supply shock and x_j the supply of good j. The term in $X^{-\psi}$ is a negative congestion externality, where X is aggregate real output (defined below) and $\psi \geq 0$. This will

[4] See Weitzman (1985), Blanchard and Kiyotaki (1987).

allow me to pick the value of ψ so as to focus on a special case which is computationally much simpler, while what is lost by doing so is independent of the point being illustrated here.

For simplicity, the aggregate price level that deflates money holdings in the utility function is assumed to be equal to the price index that is dual to the aggregate consumption index

$$c_j = \left(\int_0^1 c_{ij}^{\alpha} di \right)^{\frac{1}{\alpha}}$$

:

$$p = \left[\int_0^1 p_i^{-\frac{\alpha}{1-\alpha}} di \right]^{-\frac{1-\alpha}{\alpha}} .$$

The usual derivations concerning demand functions and aggregation are made in the Appendix. We can show that each yeoman farmer maximizes the indirect utility function given by:

$$E \ln \left[p_j^{-\frac{\alpha}{1-\alpha}} - \phi_j p_j^{-\frac{1+\mu}{1-\alpha}} \right], \tag{2}$$

where ϕ_j is a composite shock defined by

$$\phi_j = z_j M^{\mu+\psi} p^{1-\psi+\frac{\alpha\mu}{1-\alpha}} \tag{3}$$

is a composite shock and the second term is treated as constant by the agent since it is independent of his pricing policy.

From now on, I will assume that ψ is such that the composite shock does not depend on the aggregate price level: $\psi = 1 + \alpha\mu/(1-\alpha)$. Thus, $\phi_j = z_j M^{\frac{\mu+1-\alpha}{1-\alpha}}$; spillovers in price formation across firms are shut down, which greatly simplifies the analysis. It is then useful do define γ as $\gamma = \frac{\mu+1-\alpha}{1-\alpha}$.

As a benchmark, we can derive the flexible price equilibrium with no informational constraint where a different price is set for each realization of

ϕ_j. The FOC for price-setting is equivalent to

$$\begin{aligned} p_j &= \left(\frac{(1+\mu)\phi_j}{\alpha} \right)^{1/\gamma} \\ &= \left(\frac{1+\mu}{\alpha} \right)^{1/\gamma} z_j^{1/\gamma} M. \end{aligned} \tag{4}$$

Integrating we get the aggregate price level:

$$p = M \tilde{z}^{1/\gamma} \left(\frac{1+\mu}{\alpha} \right)^{1/\gamma},$$

where $\tilde{z}$ is an aggregate of z defined as

$$\tilde{z} = \left[\int_0^1 z_j^{-\frac{\alpha}{1-\alpha+\mu}} dj \right]^{-\frac{1-\alpha+\mu}{\alpha}}.$$

Thus money is neutral, the aggregate price level is proportional to M, and real aggregate output is constant and equal to

$$X = \frac{Y}{p} = \frac{M}{p} = \left(\tilde{z} \frac{1+\mu}{\alpha} \right)^{-1/\gamma}.$$

Output is lower, the larger the aggregate cost index $\tilde{z}$, the larger the elasticity of the disutility of effort μ, and the lower the elasticity of demand for the individual goods, i.e. the larger the markup over marginal cost $1/\alpha$.

The New Keynesian literature takes this framework and imposes some nominal price rigidity. I now introduce capacity constraints in processing information along the lines discussed above and derive the associated behaviour of output and the price level.

Under rational inattention, people do not have the information processing ability to pursue a rule like (4) for any value of ϕ_j. Instead they are going to pursue a rule such that the mutual information between p_j and ϕ_j satisfies a capacity constraint. Let us assume that, as in the above analysis, they pursue a discrete deterministic rule and partition the support of ϕ_j into intervals $I_k = [\bar{\phi}_k, \bar{\phi}_{k+1}]$ such that a constant value of p_j, denoted by $\bar{p}_k$, is pursued within each interval. We assume k varies between 0 and $N+1$, with $\bar{\phi}_0 = 0$ and $\bar{\phi}_{N+1} = +\infty$.

The distribution of the composite shock ϕ has density

$$g(\phi) = \int_0^{+\infty} f(M)M^{-\gamma}h(\phi M^{-\gamma})dM. \tag{5}$$

Individuals select the number of intervals, their bounds and their associated price levels so as to maximize:

$$\max_{N,\{\bar{\phi}_k,k=1,...,N\},\{\bar{p}_k,k=0,...,N\}} U = \sum_{k=0}^{N} \int_{\bar{\phi}_k}^{\bar{\phi}_{k+1}} g(\phi)\ln\left[\bar{p}_k^{-\frac{\alpha}{1-\alpha}} - \phi\bar{p}_k^{-\frac{1+\mu}{1-\alpha}}\right]d\phi, \tag{6}$$

subject to the information capacity constraint

$$-\sum_{k=0}^{N}\left(\int_{\bar{\phi}_k}^{\bar{\phi}_{k+1}} g(\phi)d\phi\right)\ln\left(\int_{\bar{\phi}_k}^{\bar{\phi}_{k+1}} g(\phi)d\phi\right) \leq K. \tag{7}$$

An equilibrium is therefore a set $\{N, \{\bar{\phi}_k, k = 1, ..., N\}, \{\bar{p}_k, k = 0, ..., N\}\}$ which maximizes (6) subject to (7). The solution to this problem then delivers the aggregate price level as a function $p(M)$ of the realization of the aggregate money stock. Given M, a price-setter j is in interval I_k iff $\bar{\phi}_k \leq z_j M^{\gamma} < \bar{\phi}_{k+1}$, which occurs with probability $H(\bar{\phi}_{k+1}M^{-\gamma}) - H(\bar{\phi}_k M^{-\gamma})$. Therefore, the aggregate price level $p(M)$ is given by

$$p(M) = \left(\sum_{k=0}^{N}\left(H(\bar{\phi}_{k+1}M^{-\gamma}) - H(\bar{\phi}_k M^{-\gamma})\right)\bar{p}_k^{-\frac{\alpha}{1-\alpha}}\right)^{-\frac{1-\alpha}{\alpha}}, \tag{8}$$

where by convention $H(+\infty) = 1$. This in turn allows to compute output $X = M/p(M)$. Note that the assumption made on ψ guarantees that the environment faced by each price-setter only depends on the exogenous variables and not on the prices set by other agents[5]. This greatly simplifies the computations.

I solve for such an equilibrium numerically, performing global optimization on all the possible partitions of the domain of ϕ into a finite number of

[5] Otherwise, the shock ϕ and its distribution $g()$ would themselves depend on the aggregate price level, and there would be no closed-form formula such as (8) for the latter–one would then need to search for a fixed point equilibrium rather than just an optimum.

intervals which match the informational capacity constraint. To keep things tractable the possible values for the jump points have been discretized[6].

Table 1 reports some summary statistics for the simulations. I start from a benchmark numerical exercise where both $f()$ and $h()$ are log-normal, with $E \ln M = E \ln z = 0$ and $\mathrm{Var}(\ln M) = \mathrm{Var}(\ln z) = 1$. The other parameters were $\mu = 1$ and $\alpha = 0.5$.

I first start by simulating this economy for $K = 1.2$ and I gradually loosen the information capacity constraint by increasing K. Table 1 reports the corresponding number of clusters along with the variance of log output. Figure 1 reports the behavior of output as a function of the monetary shock M. We see that for a wide range of values of the money stock the curve is quite flat: despite the small number of clusters, heterogeneity due to idiosyncratic shocks is enough to yield near neutrality at the aggregate level, a not unusual result (Caplin and Spulber (1987), Caballero and Engel (1993), Burstein and Hellwig (2007)). The curve is tilda-shaped: at small (resp. large) values of M, most firms charge their minimum (resp. maximum) price, and an increase in M boosts output. For intermediate values, a composition effect creates a force in the opposite direction, as some firms move to a cluster with a higher price. This composition effect creates a zone where money growth is contractionary, which also happens in other models of price rigidity.

Figure 2 compares the flatter portion of the output curve between a low information ($K = 1.2$) and a high information ($K = 1.5$) regime. We see that output is substantially flatter in the latter case. Nevertheless, as Table 1 shows, for local increases in capacity, the variance of output may well go up.

It is also interesting to look at the distribution of individual prices. They are reported in Figures 3 (for $K = 1.2$) and 4 (for $K = 1.6$). The dimension

[6] More precisely, there are $\bar{N}$ possible values of $\bar{\phi}_k$ separated by a probability weight of $1/(\bar{N}+1)$, i.e. if those eligible critical values are denoted by $\tilde{\phi}_j$, $j = 1, ..., \bar{N}$, then $\int_{\tilde{\phi}_j}^{\tilde{\phi}_{j+1}} g(\phi) d\phi = 1/(\bar{N}+1)$.
In the simulations, one has picked $\bar{N} = 20$.

of each rectangle along the y-axis is the price and along the x-axis it is the probability weight associated with the corresponding interval of values of ϕ. We see that the probability weights on each price are decreasing with the price, meaning that price-setters are devoting more attention to situations where the required price is higher. This is presumably due to the marginal disutility of labor schedule: the utility cost of not paying attention to these states is high because if one charges too low a price the labor input must be very high[7].

Entropy	#of clusters	Variance of output
1.2	4	0.17
1.3	5	0.18
1.4	5	0.187
1.5	5	0.1
1.6	6	0.1

Table 1.

Table 2 analyses the effect of an increase in the variance of monetary shocks on the distribution of individual prices for $K = 1.4$. We compare the benchmark situation to one such that $\mathrm{Var}(\ln M) = 2$ and $E(\ln M) = -0.5$ (Implying that $E(M)$ is the same as in the benchmark). We see that the increase in the variance of money shocks compells price-setters to devote more attention to high realization of those shocks[8]: the upper-tail of the distribution of the composite shocks is split into more, and finer, clusters, while the first interval is coarser. Also, the variance of log output increases from 0.19 to 0.49.

Table 3c performs the reverse exercise of dividing the variance of monetary shocks by 2, while again adjusting the mean log of M to hold $E(M)$ constant.

[7] This clearly rests on my assumption that demand must be met; this might not remain realistic for very high realizations of the demand shock.

[8] That is because of the skewness of the log-normal distribution along with the increasing marginal disutility of labor property. But for even larger increases in the variance of money shocks, the price setters will also spend information capacity on the lower tail of the distribution. Thus, for $\mathrm{Var}(\ln M) = 4$, cluster 1 has a minimal weight of 0.048.

We see that the number of clusters is the same, and so is their size, but the order is changed: the second cluster gets the biggest weight, while more attention is paid to low realizations of the shock than before. The intuition for this result is unclear.

Cluster	Price	Weight
1	0.95	0.43
2	1.96	0.24
3	3.74	0.19
4	7.0	0.095
5	23.63	0.048

Table 2a – $K = 1.4$, benchmark.

Cluster	Price	Weight
1	0.76	0.52
2	1.67	0.19
3	2.84	0.095
4	4.99	0.095
5	8.66	0.048
6	47.4	0.048

Table 2b – $K = 1.4$, $\text{Var}(\ln M) = 2$ and $E(\ln M) = -0.5$.

Cluster	Price	Weight
1	0.81	0.19
2	1.93	0.43
3	3.65	0.24
4	6.02	0.095
5	13.45	0.048

Table 2c – $K = 1.4$, $\text{Var}(\ln M) = 0.5$ and $E(\ln M) = 0.25$

5 Implications for rigidity

As the preceding section makes clear, the quantized model only implies a moderate degree of price rigidity at the aggregate level. By contrast, in a model where noisy behavior is allowed for, as that of Mackoviak and Wiederholt (2009a,b), the aggregate price level reacts less than one for one to monetary shocks throughout the whole distribution of those shocks. The reason is that this class of models is similar to the Lucas (1972, 1973) misperception model. Information capacity constraints preclude price-setters from reacting to the monetary shock. Instead, they can only react to a noisy signal of the monetary shock. Given that, their optimal inference about the true realization of the money stock will react less than one-for-one to that money stock, as implied by Bayes's Law. The only difference between the Lucas misperception model is that the noise is now designed optimally by the price-setters so as to meet the information capacity constraint.[9] In the aggregate, all individual prices react less than one-for-one to the money shock, and so does the aggregate price level. These considerations apply to any model where

[9] As an aside, it is interesting to note that the reaction of prices to monetary shocks is optimal conditional on the existence of noise. In other words, it is not the information capacity constraint which is constraining that reaction to be suboptimally low, but rather the noise itself (of course the noise is also a by-product of the information capacity constraint). To see this, let us get back to the standard Gaussian linear-quadratic problem:

$$V = \min E(x-y)^2.$$

Assume $y \sim N(0, \sigma_y^2)$ and $x = ay + \varepsilon$, where a is a reaction coefficient and ε the endogenous noise, assumed normal with zero mean and variance σ_ε^2 and orthogonal to y.

An optimality condition for x is

$$E(y \mid x) = x. \tag{9}$$

This optimality condition pins down the correlation between x and y and it would hold if one observed an exdogenous noisy signal of y. In our context, we have $E(y \mid x) = \frac{a\sigma_y^2}{\sigma_\varepsilon^2 + a^2\sigma_y^2}$. Therefore the optimality condition (9) is equivalent to

$$a(1-a) = \frac{\sigma_\varepsilon^2}{\sigma_y^2}. \tag{10}$$

The value of the objective function is $V = (a-1)^2\sigma_y^2 + \sigma_\varepsilon^2$; The mutual information between x and y is

$$M(x,y) = \frac{1}{2}(\log(a^2\sigma_y^2 + \sigma_\varepsilon^2) - \log(\sigma_\varepsilon^2)). \tag{11}$$

Thus our problem is equivalent to maximizing V subject to

$$\frac{a^2\sigma_y^2}{\sigma_\varepsilon^2} \leq K. \tag{12}$$

Given this constraint, which involves a, it is not a priori obvious that the optimality condition (9) should hold. In contrast, if x and y have a discrete distribution, $M(x,y)$ only depend on the probability weights of their joint distribution, and it is always possible to pick the values of x while leaving $M(x,y)$ unchanged so as to match (9). That $M(x,y)$ is not independent of the values of x because of the presence of a in (11) is a weakness of the entropy concept applied to continuous distributions.

Nevertheless, since one picks both σ_ε and a optimally given the constraint (12), one has one degree of freedom left to match the optimality condition (9)-(10), which turns out to hold at the optimum. Indeed, at the optimum $a = \frac{K}{K+1}$ and $\sigma_\varepsilon^2 = \frac{K}{(K+1)^2}\sigma_\varepsilon^2$, implying that (10) holds.

This proves that the underreaction of x to y does not come from a failure of (9) that would be the price to pay for matching the information capacity constraint. If this were the case, it would be an artifact of the use of continuous entropy. Instead, this underreaction is optimal given the presence of (endogenous) noise.

agents are allowed to introduce noise in their policy functions in order to save on information capacity. In particular, underreaction would also arise if the distribution of the exogenous variable (y) were purely discrete or if that of the endogenous variable (x) turned out to be discrete yet noisy conditional on y as in Matejka and Sims (2010).

On the other hand, in the quantized model developed here, noisy policies are precluded. The information capacity constraint is matched by lumping the realizations of y in clusters within which the same policy is pursued. The actual value of x within each cluster does not affect the mutual information between x and y, since, as we have seen, it only depends on the probability weights of the discrete random variable x. Therefore, within each cluster one will pick the optimal value of x conditional on being in that cluster, ignoring the information capacity constraint. Consequently, if, in the absence of information constraints, it is optimal for x to react one-for-one to y, this will remain so in the quantized solution when one moves across clusters. In the aggregate, money neutrality tends to arise in a similar fashion as in Caplin and Spulber (1987): the large price adjustment of firms that are near the frontier between two clusters tends to offset the price inertia of those firms that remain in the same cluster. Thus the model, resembles a menu cost model rather than the Lucas misperception model, and has much less aggregate price stickiness.

6 Discussion

The general message of this paper is that information processing constraints yield lumpy behavior. Thus, when the exogenous variables change, inattention results in inaction, while in the standard approach it is associated with inadequacy, i.e. embodies excess noise. In both cases, the endogenous variable does not react enough to the exogenous one, although here there will be a jump if one crosses the frontier between clusters.

The existence of lumpiness in the adjustment of economic variables has been documented in a number of areas. For example, Doms and Dunne (1998), studying investment at the plant level, find that "Many plants occasionally alter their capital stocks in lumpy fashions. Of the plants in a balanced panel, over half experience a capital adjustment of at least 37 % in one year, and by 50% in two consecutive years". In the area of price setting, Klenow and Kryvstov (2008) find (table III) that individual price changes are usually large, with a mean size of 14 %. Dhyne et al. (2006) report similar findings, along with substantial heterogeneity in the degree of lumpiness of price adjustment across sectors. More recently, Eichenbaum et al. (2008), using scanner data from the retail trade sector, find that firms pick their prices among a number of finite "reference" prices, and that the evidence is consistent with the view that it is not costly to change prices as long as the new price remains a reference price. This is exactly what happens in the model described above. On the other hand, reference prices change infrequently, which may be interpreted as the outcome of a costly reoptimization process in light of perceived changes in the underlying distribution of shocks or in the technology for processing information.

Finally, evidence of lumpiness in employment can be found in Davis et al (1996) or Caballero et al. (1997). The latter, in particular, found that the distribution of employment changes is typically bimodal.

Of course, rational inattention is not the only reason why there could be lumpiness. The above literature has mostly focused on fixed and linear adjustment costs and rational inattention and adjustment costs are not mutually exclusive mechanisms. The rational inattention mechanism may be of particular interest when large adjustment costs are implausible, as in the area of price setting. Furthermore, a range of novel predictions may be generated regarding the determinants of lumpiness: The greater an economic agent's ability to process information, the less lumpy its behaviour. Thus one may speculate that advanced in information technologies have reduced lumpiness[10], or that firms with a greater fraction of highly skilled workers

[10] This is the message of the empirical study by Bartel et al. (2005) inthe space of

have less lumpy behavior – this may help explain, for example, the finding by Doms and Dunne (1998) that smaller plants have a more lumpy adjustement, if one is willing to believe that smaller plants employ fewer skilled workers, or by Dhyne et al (2006, fig. 1) that some sectors (like gasoline) have much less lumpy price adjustment than others (like haircuts).

---

product diversity.

REFERENCES

Bartel, Ann, Ann P., Casey Ichniowski and Kathryn L. Shaw (2005), "How Does Information Technology Really Affect Productivity? Plant-Level Comparisons of Product Innovation, Process Improvement and Worker Skills"NBER WP #11773.

Blanchard, Olivier, and Nobuhiro Kiyotaki (1987), "Monopolistic Competition and the Effects of Aggregate Demand Monopolistic Competition and the Effects of Aggregate Demand", *American Economic Review*, 77, 4, 647-666

Burstein, Ariel and Christian Hellwig (2007), "Prices and Market Shares in a Menu Cost Model", NBER working paper.

Caballero, Ricardo and Eduardo M. R. A. Engel (1993), "Heterogeneity and Output Fluctuations in a Dynamic Menu-Cost Economy" *Review of Economic Studies*, 60, 1, 95-119.

Caballero, Ricardo, Engel, Eduardo and John Haltiwanger (1997), "Aggregate Employment Dynamics: Building from Microeconomic Evidence", *American Economic Review,* 87, 1, 115-137

Caplin, Andrew S. and Daniel F. Spulber (1987), "Menu Costs and the Neutrality of Money" *Quarterly Journal of Economics*, 102, 4, 703-726

Davis, Steven J., John Haltiwanger and Scott Schuh (1996), *Job creation and destruction.* Cambridge MA: MIT Press

Dhyne, Emmanuel, Luis J. Alvarez, Hervé Le Bihan, Giovanni Veronese, Daniel Dias, Johannes Hoffmann, Nicole Jonker, Patrick Lünnemann, Fabio Rumler and Jouko Vilmunen (2006), "Price Changes in the Euro Area and the United States: Some Facts from Individual Consumer Price Data" *Journal of Economic Perspectives,* 20, 2, 171-192

Doms, Mark E. and Timothy Dunne (1998), "Capital Adjustment Patterns in Manufacturing Plants" *Review of Economic Dynamics*, 1(2), 409-429.

Eichenbaum, Martin, Nir Jaimovich and Sergio Rebelo (2008), "Reference Prices and Nominal Rigidities", NBER Working Paper 13829.

Klenow, Peter and Oleksiy Kryvtsov (2008), "State-dependent or timedependent pricing: does it matter for recent US inflation?", *Quarterly Journal of Economics,* 123, 3, 863-904

Lucas, Robert E. (1972), "Expectations and the neutrality of money", *Journal of Economic Theory*

——————— (1973), "International evidence on the output-inflation trade-off", *American Economic Review*

Luo, Yulei (2008): "Consumption Dynamics under Information Processing Constraints."

Review of Economic Dynamics, 11: 366-385.

Mackowiak, Bartosz and Mirko Wiederholt (2009a): "Optimal Sticky Prices under Rational

Inattention." *American Economic Review*, 99, 769-803.

Mackowiak, Bartosz and Mirko Wiederholt (2009b), "Business Cycle Dynamics under Rational Inattention", mimeo, ECB.

Mankiw, G., Reis, R., 2002. "Sticky information versus sticky prices: A proposal to replace the New Keynesian Phillips curve", *Quarterly Journal of Economics* 117 (4), 1295-1328.

Matejka, Filip (2008), "Rigid Pricing and rationally Inattentive Consumer", Working Paper, Princeton University

Matejka, Filip and Christopher A. Sims (2010), "Discrete Actions in Information Constrained Tracking Problems", Working Paper, Princeton University

Moscarini, G., (2004) "Limited information capacity as a source of inertia", *Journal of Economic Dynamics and Control* 28(10), 2003-2035.

Paciello, Luigi (2007), "The Response of Prices to Technology and Monetary Policy Shocks under Rational Inattention", mimeo, Einaudi Institute for Economics and Finance

Saint-Paul, Gilles (2005) "Some evolutionary foundations for price level rigidity", *American Economic Review.*

Shannon, Claude (1948), "A Mathematical Theory of Communication", *Bell System Technical Journal*, vol. 27, p. 379-423

Sims, Christopher A. (2003): "Implications of Rational Inattention." *Journal of Monetary*

Economics, 50(3), 665-690.

Sims, Christopher A. (2006): "Rational Inattention: Beyond the Linear-Quadratic

Case." *American Economic Review*, 96(2): 158-163.

Weitzman, Martin L. "The Simple Macroeconomics of Profit Sharing", *American Economic Review*, 75, 5 (1985), 937-953

APPENDIX

Proof of Theorem 1.

We have that

$$M(\varepsilon) = H(Y;\varepsilon) - H_X(Y;\varepsilon).$$

Let us compute $H_x(y;\varepsilon)$. We denote by $p(y)$ the distribution of y and by $g(x) = \int_y f_\varepsilon(x \mid y)p(y)dy$ the unconditional distribution of x. By Bayes's law the conditional distribution of y is $f_\varepsilon(y \mid x) = \frac{f_\varepsilon(x|y)p(y)}{g(x)}$. Therefore we have that

$$\begin{aligned}
H_x(Y;\varepsilon) &= \int_x \left(- \int_y \frac{f_\varepsilon(x \mid y)p(y)}{g(x)} \log \frac{f_\varepsilon(x \mid y)p(y)}{g(x)} \right) g(x)dx \\
&= \int_x \left(- \int_y f_\varepsilon(x \mid y)p(y) \log \frac{f_\varepsilon(x \mid y)p(y)}{g(x)} \right) dx \\
&= \int_x \Bigg(- \int_y \frac{1}{\varepsilon}\hat{f}\left(\frac{x - x(y)}{\varepsilon}\right) p(y) \Bigg[\log\left(\frac{1}{\varepsilon}\hat{f}\left(\frac{x - x(y)}{\varepsilon}\right)p(y)\right) - \log\Bigg(\int_u \frac{1}{\varepsilon}\hat{f}\Bigg(\frac{x - x(u}{\varepsilon} \\
&= I_1 + I_2,
\end{aligned}$$

where

$$I_1 = - \int_x \left(\int_y \frac{1}{\varepsilon}\hat{f}\left(\frac{x - x(y)}{\varepsilon}\right) p(y) \log\left(\frac{1}{\varepsilon}\hat{f}\left(\frac{x - x(y)}{\varepsilon}\right) p(y)\right) dy \right) dx$$

and

$$I_2 = \int_x \left(\int_y \frac{1}{\varepsilon}\hat{f}\left(\frac{x - x(y)}{\varepsilon}\right) p(y)dy \right) \log\left(\int_u \frac{1}{\varepsilon}\hat{f}\left(\frac{x - x(u)}{\varepsilon}\right) p(u)du \right) dx.$$

Let $H_n(Y) = - \int_{T_n} p(y)\log p(y)dy$. Clearly, $\sum_n H_n(Y) = H(Y)$. Furthermore,

$$\begin{aligned}
\int_y \frac{1}{\varepsilon}\hat{f}\left(\frac{x - x(y)}{\varepsilon}\right) p(y) \log\left(\frac{1}{\varepsilon}\hat{f}\left(\frac{x - x(y)}{\varepsilon}\right) p(y)\right) dy &= \sum_n \int_{S_n} \frac{1}{\varepsilon}\hat{f}\left(\frac{x - x_n}{\varepsilon}\right) p(y) \log\Bigg(\frac{1}{\varepsilon}\hat{f}\Bigg(\\
&= \sum_n \frac{1}{\varepsilon}\hat{f}\left(\frac{x - x_n}{\varepsilon}\right) (\log\left(\frac{1}{\varepsilon}\hat{f}\left(\frac{x - x_n}{\varepsilon}\right)\right.
\end{aligned}$$

Therefore,

$$\begin{aligned}
I_1 &= -\int_x \left[\sum_n \frac{1}{\varepsilon}\hat{f}\left(\frac{x-x_n}{\varepsilon}\right)\left(\log\left(\frac{1}{\varepsilon}\hat{f}\left(\frac{x-x_n}{\varepsilon}\right)\right)P_n + H_n(Y)\right)\right]dx \\
&= -\sum_n P_n \int_x \frac{1}{\varepsilon}\hat{f}\left(\frac{x-x_n}{\varepsilon}\right)\log\left(\frac{1}{\varepsilon}\hat{f}\left(\frac{x-x_n}{\varepsilon}\right)\right)dx - \sum_n H_n(Y)\int_x \frac{1}{\varepsilon}\hat{f}\left(\frac{x-x_n}{\varepsilon}\right)dx \\
&= \log\varepsilon + H_{\hat{f}} + H(Y),
\end{aligned}$$

where $H_{\hat{f}} = -\int_z \hat{f}(z)\log\hat{f}(z)dz$ is the entropy of distribution $\hat{f}()$ and the last equality can be obtained straightforwardly by considering the variable change $z = \frac{x-x_n}{\varepsilon}$.

Next, we have that

$$\int_y \frac{1}{\varepsilon}\hat{f}\left(\frac{x-x(y)}{\varepsilon}\right)p(y)dy = \sum_n \frac{1}{\varepsilon}\hat{f}\left(\frac{x-x_n}{\varepsilon}\right)P_n.$$

Therefore,

$$\begin{aligned}
I_2 &= \int_x \sum_n \frac{1}{\varepsilon}\hat{f}\left(\frac{x-x_n}{\varepsilon}\right)P_n \log\left(\sum_k \frac{1}{\varepsilon}\hat{f}\left(\frac{x-x_k}{\varepsilon}\right)P_k\right)dx \\
&= -\log\varepsilon + \sum_n P_n \int_x \frac{1}{\varepsilon}\hat{f}\left(\frac{x-x_n}{\varepsilon}\right)\log\left(\sum_k \hat{f}\left(\frac{x-x_k}{\varepsilon}\right)P_k\right)dx \\
&= -\log\varepsilon + \sum_n P_n \int_z \hat{f}(z)\log\left(\sum_k \hat{f}\left(z+\frac{x_n-x_k}{\varepsilon}\right)P_k\right)dz.
\end{aligned}$$

It is then easy to see that since $\hat{f}()$ is regular and since $\lim_{|z|\to+\infty}\hat{f}(z)=0$, over any bounded interval $[a,b]$ the family of functions $v_\varepsilon()$ defined by $v_\varepsilon(z) = \hat{f}(\varepsilon)\log\left(\sum_k \hat{f}\left(z+\frac{x_n-x_k}{\varepsilon}\right)P_k\right)$ converges uniformly to $\hat{f}(z)\log\left(\hat{f}(z)P_n\right)$. Furthermore since for any z, $\log\left(\hat{f}(z)P_n\right) \le \log\left(\sum_k \hat{f}\left(z+\frac{x_n-x_k}{\varepsilon}\right)P_k\right) \le \log\max_z \hat{f}(z) = \mu$, for any set S we have

$$\int_S \hat{f}(z)\log\hat{f}(z) + (\log P_n)\int_S \hat{f}(z)dz \le \int_S \hat{f}(z)\log\left(\sum_k \hat{f}\left(z+\frac{x_n-x_k}{\varepsilon}\right)P_k\right)dz \le \mu\int_S \hat{f}(z)dz.$$

Clearly we can always pick a and b such that for $S=\mathbb{R}-[a,b]$ the two expressions in the right and left are arbitrarily small. This must also be true of the expression in the middle, for any ε. By uniform convergence, we then know that for ε small enough $\int_a^b \hat{f}(z) \log\left(\sum_k \hat{f}\left(z+\frac{x_n-x_k}{\varepsilon}\right) P_k\right) dz$ is arbitrarily close to $\int_a^b \hat{f}(z) \log\left(\hat{f}(z) P_n\right)$. Since both $\int_{\mathbb{R}-[a,b]} \hat{f}(z) \log\left(\hat{f}(z) P_n\right)$ (the left term in the inequality) and $\int_{\mathbb{R}-[a,b]} \hat{f}(z) \log\left(\sum_k \hat{f}\left(z+\frac{x_n-x_k}{\varepsilon}\right) P_k\right) dz$ are arbitrarily small, we have shown that $\int_{-\infty}^{+\infty} \hat{f}(z) \log\left(\sum_k \hat{f}\left(z+\frac{x_n-x_k}{\varepsilon}\right) P_k\right) dz$ is arbitrarily close to $\int_{-\infty}^{+\infty} \hat{f}(z) \log\left(\hat{f}(z) P_n\right)$.

From this we get that

$$\begin{aligned}
\lim_{\varepsilon\to 0}(I_2+\log\varepsilon) &= \lim_{\varepsilon\to 0}\sum_n P_n \int_z \hat{f}(z)\log\left(\sum_k \hat{f}\left(z+\frac{x_n-x_k}{\varepsilon}\right) P_k\right) dz \\
&= \sum_n P_n \int_{-\infty}^{+\infty} \hat{f}(z)\log\left(\hat{f}(z) P_n\right). \\
&= -H_{\hat{f}} - S(X).
\end{aligned}$$

Therefore:

$$\begin{aligned}
\lim_{\varepsilon\to 0} H_X(Y;\varepsilon) &= \lim_{\varepsilon\to 0}\log\varepsilon + H_{\hat{f}} + H(Y) + I_2 \\
&= H_{\hat{f}} + H(Y) - H_{\hat{f}} - S(X) \\
&= H(Y) - S(X).
\end{aligned}$$

Consequently, $M(\varepsilon) = H(Y) - H_X(Y;\varepsilon)$ converges to $S(X)$ as $\varepsilon \to 0$. QED.

Proof of Lemma 1.

Consider $y \in \mathring{S}_n$ and $y' \in \mathring{S}_p$ for $p \neq n$. A swap between y and y' would be marginally improving iff

$$(x_n - y')^2 + (x_p - y)^2 < (x_n - y)^2 + (x_p - y')^2,$$

or equivalently

$$(x_n - x_p)(y - y') > 0.$$

Let us now assume that for some n, $\partial\mathring{S}_n$ is not convex. Then we can find $y, y'' \in C(\mathring{S}_n)$, and $\lambda \in (0,1)$ such that $y' = \lambda y + (1-\lambda)y'' \notin C(\mathring{S}_n)$. Since y

and y'' are limits of sequences of elements of $\mathring{S}_n$, we can simply assume they are themselves interior. Since $y' \notin C(\mathring{S}_n)$ and $C(\mathring{S}_n)$ is closed, there exists an interval $I' = (\lambda - \varepsilon, \lambda + \varepsilon)$ such that $y'(\lambda') = \lambda' y + (1 - \lambda')y'' \notin C(\mathring{S}_n)$ for all $\lambda' \in (\lambda - \varepsilon, \lambda + \varepsilon)$. Since $y, y'' \in \mathring{S}_n$, we can construct two intervals $J = (y - \eta, y + \eta) \subset \mathring{S}_n$ and $J'' = (y'' - \eta, y'' + \eta) \subset \mathring{S}_n$ and furthermore choose them such that they do not intersect $J' = y'(I')$. Finally there exists a mapping $y()$ from I' to J and $y''()$ from I' to J'' which allows us to index those two sets by λ'.

For $\lambda' \in I'$ $(y'(\lambda') - y(\lambda'))(y'(\lambda') - y''(\lambda')) < 0$. Thus, either $(x_n - x(y'(\lambda')))(y(\lambda') - y'(\lambda')) > 0$ or $(x_n - x(y'(\lambda')))(y''(\lambda') - y'(\lambda')) > 0$. Let $y^*(\lambda')$ be the element of $\{y(\lambda'), y''(\lambda')\}$ such that the inequality is satisfied. Consider now the new assignment which, for all the λ's, replaces $x(y'(\lambda'))$ with x_n and sets $x(y^*(\lambda')) = x(y'(\lambda'))$ instead of x_n. Each of those swaps is marginally improving the objective function by an amount which is bounded from below by a strictly positive number, since there is a finite number of values of x_n and $\mid y'(\lambda') - y^*(\lambda') \mid$ is also bounded away from zero. Integrating those gains over I' and noting that $y()$ has full support, we see that the objective function must improve by a strictly positive amount, which contradicts the initial optimality condition. Thus $C(\mathring{S}_n)$ must be convex. QED.

Proof of Theorem 2.

Lemma 1 implies that any optimum must be a partition by intervals, up to a set of measure zero. It follows that one cannot improve on such a partition. By Lemma 2, for any partition the optimal x_n must be $\frac{y_n + y_{n+1}}{2}$, which proves (v). Next, computing the value function for such a configuration, we get that

$$E(x(y) - y)^2 = \frac{1}{12} \sum_{n=0}^{N-1} (y_{n+1} - y_n)^3. \tag{13}$$

For a given N, we minimize (13) subject to

$$\begin{aligned} y_0 &= 0, \\ y_N &= 1, \\ -\sum_{n=0}^{N-1}(y_{n+1}-y_n)\ln(y_{n+1}-y_n) &\leq K. \end{aligned}$$

The FOCs are:

$$(y_n-y_{n-1})^2-(y_{n+1}-y_n)^2 = \lambda(\ln(y_n-y_{n-1})-\ln(y_{n+1}-y_n)),\ 0<n<N. \quad (14)$$

Note that absent the capacity constraint, optimality would imply that $y_n - y_{n-1} = y_{n+1} - y_n$. All intervals would then be of constant length $1/N$ and the resulting entropy would be $\ln N$. Thus, if $\ln N < K$, then $\lambda = 0$ and the optimal solution is the unconstrained one. However, one can always improve on this by picking a larger N, since the initial configuration can always be replicated by collapsing the additional interval to a set of measure zero by equating their bounds. Therefore the optimal N will be such that $\ln N \geq K$, i.e. the capacity constraint will be binding. Let us then consider such an N. Call Δ_n the length of interval n. The FOC (14) implies that $\Delta_n^2 - \lambda \ln \Delta_n$ is invariant across intervals. Since the function $X^2 - \lambda \ln X$ is U-shaped, Δ_n can at most have two values, let us call them Δ and Δ'. Clearly, the invariance property is then satisfied for $\lambda = \frac{\Delta'^2-\Delta^2}{\ln \Delta' - \ln \Delta}$. Without loss of generality, assume $\Delta \leq \Delta'$. Let q the number of intervals of length Δ. Since the whole [0,1] interval must be partitioned, it must be that

$$q\Delta + (N-q)\Delta' = 1$$

and

$$-q\Delta \ln \Delta - (N-q)\Delta' \ln \Delta' = K.$$

Eliminating Δ', we get

$$\Delta' = \frac{1-q\Delta}{N-q},$$

and we see that Δ must solve

$$\phi(\Delta) = -q\Delta \ln \Delta - (1 - q\Delta) \ln \left(\frac{1 - q\Delta}{N - q} \right) = K. \tag{15}$$

The function $\phi(\Delta)$ is increasing and then decreasing and reaches its maximum at $\Delta = 1/N$, at which point we also have $\Delta' = 1/N$. Therefore, there is at most one solution Δ such that $\Delta \leq \Delta'$. Furthermore, $\phi(0) = \ln(N - q)$ and $\phi(1/N) = \ln N$. Therefore, there exists a solution for Δ provided

$$\ln(N - q) < K \leq \ln N.$$

In particular, for any N the set of values of q for which this holds is non empty.

Despite that q is integer, equation (15) also defines a value of Δ for any real number q. Furthermore,

$$\frac{\partial \phi}{\partial q} = \Delta(\ln \Delta' - \ln \Delta) + \Delta - \Delta' < 0.$$[11]

Since $\phi'(\Delta) > 0$, it follows that $\frac{d\Delta}{dq} > 0$.

Next, note that the resulting loss function, up to a positive multiplicative constant, is equal to $V = q\Delta^3 + (N - q)\Delta'^3$. Differentiating, we get

$$\begin{aligned} dV &= (\Delta^3 - \Delta'^3)dq - 3q\Delta'^2 dq + 3\Delta'^3 dq + 3q\Delta^2 d\Delta - 3q\Delta'^2 d\Delta \\ &= (\Delta^3 - \Delta'^3)dq + 3\Delta'^2(\Delta' - q)dq + 3q(\Delta^2 - \Delta'^2)d\Delta. \end{aligned}$$

Since $\frac{d\Delta}{dq} > 0$, $\Delta < \Delta'$, and $\Delta' < 1 \leq q$, all terms are negative if $dq > 0$. Therefore, V is a decreasing function of q; given N, the optimal value of q is the largest possible one, i.e. $q = N - 1$. The resulting loss function is then

$$V = (N - 1)\Delta^3 + (1 - (N - 1)\Delta)^3, \tag{16}$$

and Δ now solves

$$\tilde{\phi}(\Delta) = -(N - 1)\Delta \ln \Delta - (1 - (N - 1)\Delta) \ln (1 - (N - 1)\Delta) = K. \tag{17}$$

[11] It can be checked that this expression is always negative by noting that it would be equal to zero at $\Delta = \Delta'$ and that its derivative with respect to Δ' is $\Delta/\Delta' - 1 < 0$.

What is the optimal value of N? First of all, differentiating $\tilde{\phi}$ with respect to N and Δ we get

$$\frac{d\Delta}{dN} = -\frac{\Delta}{N-1}(1 + \frac{1}{\ln\Delta' - \ln\Delta}) < 0. \tag{18}$$

Next, differentiating (16) and using (18) we get that

$$\frac{dV}{dN} = \Delta^3 - 3\Delta\Delta'^2 + 3\Delta(\Delta'^2 - \Delta^2)(1 + \frac{1}{\ln\Delta' - \ln\Delta}).$$

This expression is positive if and only if

$$2\Delta^2 < \frac{3(\Delta + \Delta')(\Delta' - \Delta)}{\ln\Delta' - \ln\Delta}.$$

Calling $\theta = \Delta'/\Delta > 1$, this is equivalent to $\ln\theta < 3(\theta^2 - 1)/2$, which is always true.

Thus $dV/dN > 0$. Consequently, the optimal value of N is the smallest one such that $\ln N \geq K$, i.e. $N = INT(e^K)$.

QED

Derivation of (2)-(3).

The budget constraint of the individual is

$$\int_0^1 p_i c_{ij} + m_j \leq y_j + s_j,$$

where $y_j = p_j x_j$ is his income and s_j is rebated seignioriage. In equilibrium the total money stock is $M = \int_0^1 m_j dj$ and we assume for simplicity that seignoriage is rebated proportionally to the value of output produced by the individual:

$$s_j = M\frac{y_j}{Y}, \forall j,$$

where

$$Y = \int_0^1 y_j dj$$

is GDP. Aggregate real output is defined as $X = \left(\int_0^1 x_j^\alpha dj\right)^{1/\alpha}$.

We assume that the money stock is drawn from a distribution with density $f(M)$ and c.d.f $F(M)$. We also assume that the idiosyncratic shock is drawn from a distribution with density $h(z)$ and cumulative $H(z)$.

Solving for the consumer's optimal consumption and money holdings yields, after a few steps, the following relationship:

$$c_{ij} = \frac{m_j}{p_i^{\frac{1}{1-\alpha}} p^{-\frac{\alpha}{1-\alpha}}}. \tag{19}$$

Aggregating across individuals, this gives the demand curve for good i :

$$C_i = \frac{M}{p_i^{\frac{1}{1-\alpha}} p^{-\frac{\alpha}{1-\alpha}}}. \tag{20}$$

We assume that all producers meet demand. Therefore, $x_j = C_j$.Next,

$$\begin{aligned} Y &= \int p_j x_j dj \\ &= \int p_j C_j dj \\ &= M. \end{aligned}$$

We can also wheck that $X = \left(\int C_i^{\alpha} di\right)^{1/\alpha} = M/p$.

Furthermore, aggregating (19) across goods we see that the aggregate consumption index for individual j is equal to

$$c_j = \frac{m_j}{p}.$$

We also have that $\int_0^1 p_i c_{ij} = m_j = pc_j$. Substituting into the budget constraint, we get that

$$\begin{aligned} m_j &= \frac{y_j + s_j}{2}; \\ c_j &= \frac{y_j + s_j}{2p}. \end{aligned}$$

Noting that $s_j = M\frac{y_j}{Y}$ and $y_j = p_j x_j$ we get an indirect utility function

$$\begin{aligned} V_j &= E\ln\left[c_j^{1/2}\left(\frac{m_j}{p}\right)^{1/2}\left(\frac{M}{p}\right)^{-\psi} - z_j x_j^{1+\mu}\right] \\ &= E\ln\left[\frac{p_j x_j(1+M/Y)}{2p}\left(\frac{M}{p}\right)^{-\psi} - z_j x_j^{1+\mu}\right] \\ &= .E\ln\left[\frac{p_j x_j}{p}\left(\frac{M}{p}\right)^{-\psi} - z_j x_j^{1+\mu}\right] \end{aligned} \tag{21}$$

It is this quantity that the individual maximizes when setting his price p_j subject to the demand curve (20). Substituting this demand curve into (21) we can rewrite the objective function of the producer as

$$\begin{aligned} V_j &= E\ln\left[\left(\frac{p_j}{p}\right)^{-\frac{\alpha}{1-\alpha}}\frac{M}{p}\left(\frac{M}{p}\right)^{-\psi} - z_j\frac{M^{1+\mu}}{p_j^{\frac{1+\mu}{1-\alpha}}p^{-\frac{\alpha(1+\mu)}{1-\alpha}}}\right] \\ &= E\ln\left[p_j^{-\frac{\alpha}{1-\alpha}} - \phi_j p_j^{-\frac{1+\mu}{1-\alpha}}\right] + E\ln\left[p^{\frac{2\alpha-1}{1-\alpha}+\psi}M^{1-\psi}\right], \end{aligned}$$

where ϕ_j is defined by (3). This clearly amounts to maximizing (2).

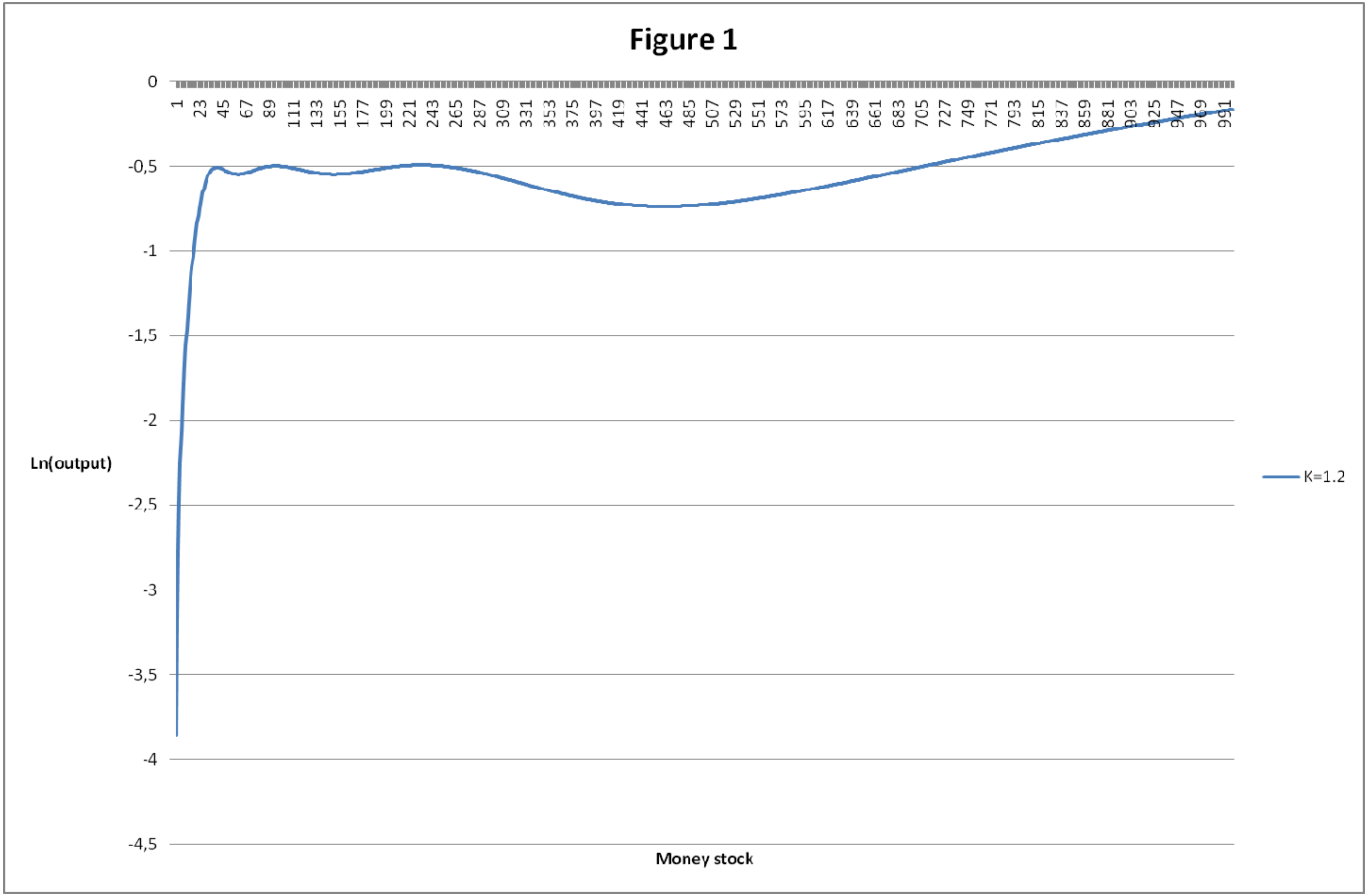
Figure 1
0
-0,5
-1
-1,5
-2
-2,5
-3
-3,5
-4
-4,5
Ln(output)
Money stock
K=1.2

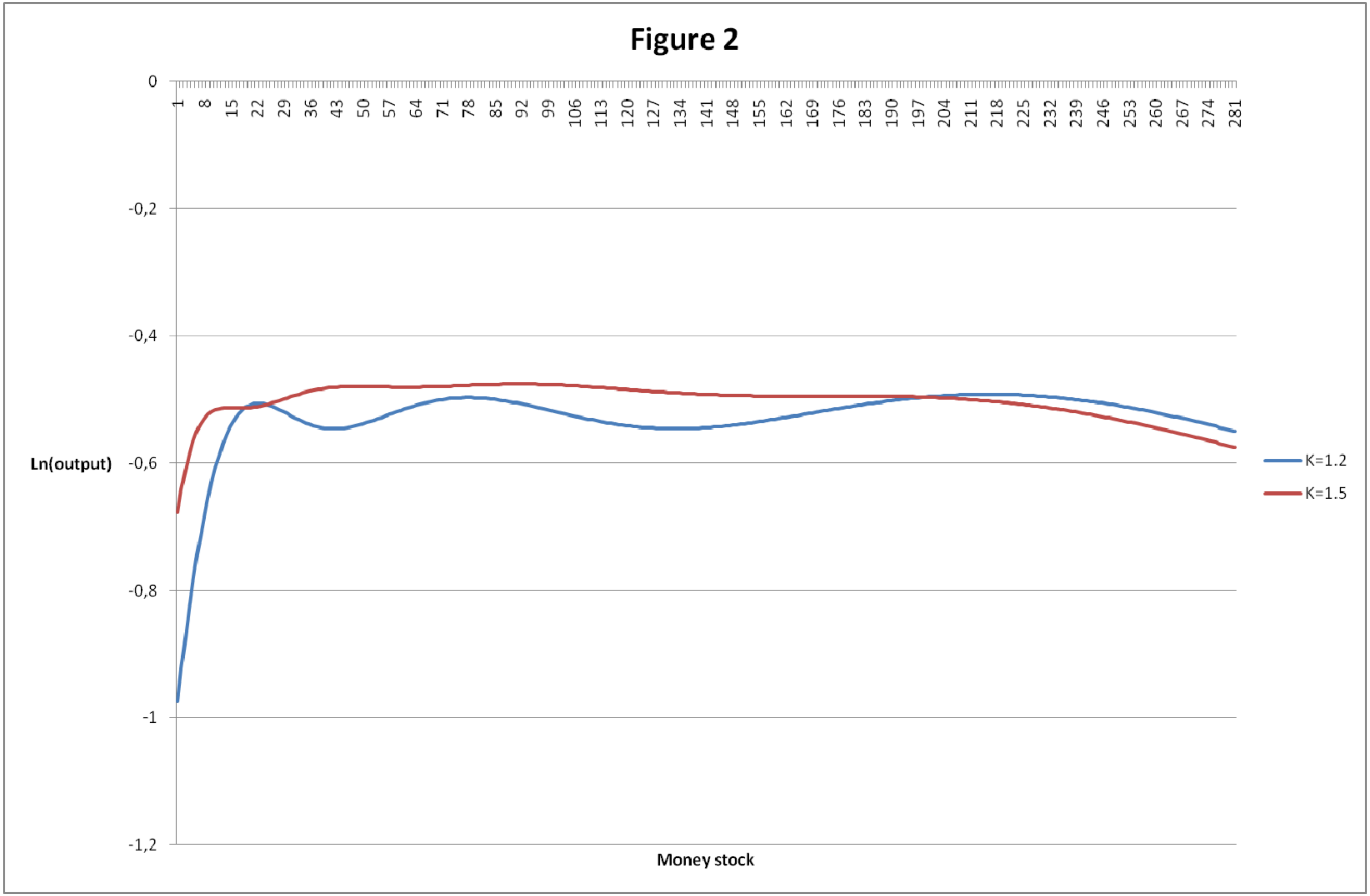
Figure 2
0
-0,2
-0,4
-0,6
-0,8
-1
-1,2
Ln(output)
Money stock
1
8
15
22
29
36
43
50
57
64
71
78
85
92
99
106
113
120
127
134
141
148
155
162
169
176
183
190
197
204
211
218
225
232
239
246
253
260
267
274
281
K=1.2
K=1.5

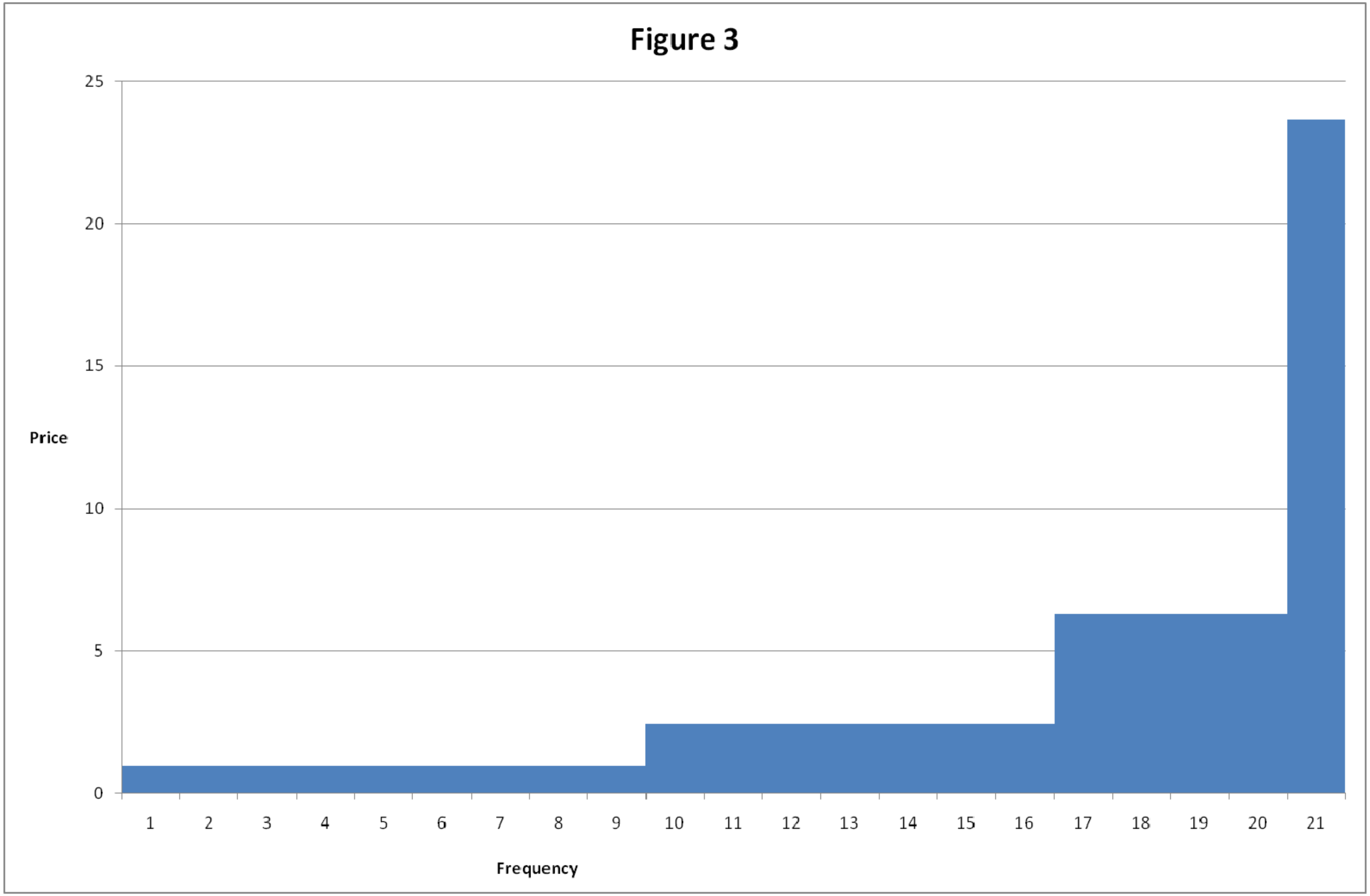
Figure 3
25
20
15
10
5
0
Price
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Frequency

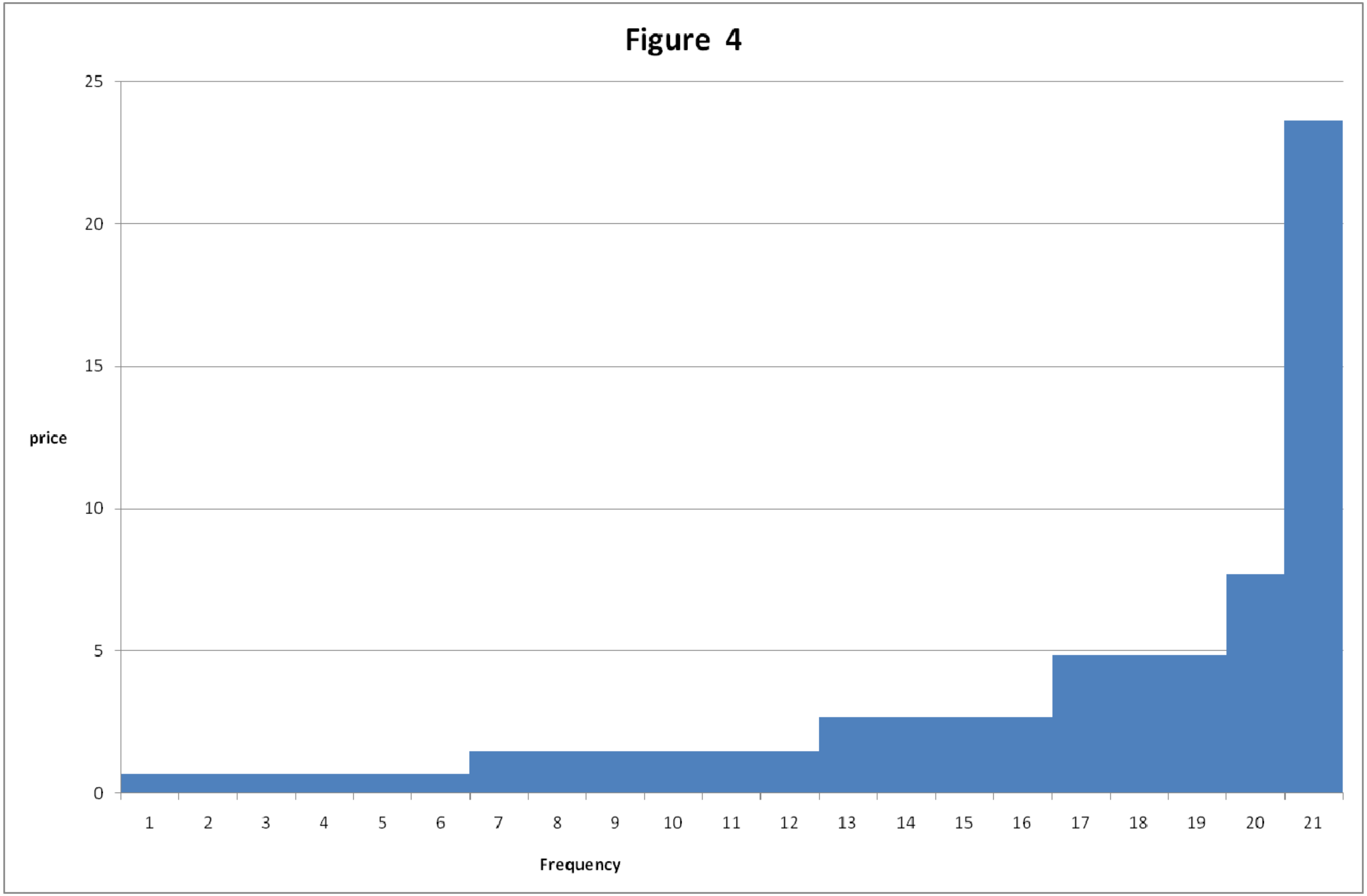
Figure 4
price
Frequency
0
5
10
15
20
25
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21