E C O N O M I C S   B U L L E T I N

A Model of Mental Effort and Endogenous Estimation Risk

Diego Nocetti
The University of Memphis

Abstract

I present a simple model that formalizes Kahneman's (1973) ideas and experimental work on attention limitations. In addition, I extend his framework to account for the interaction between attention and memory deficits. In particular, I propose that individuals optimally allocate their divisible, but limited, attention to estimate parameters of an economic model, by retrieving observations from a stock of memories, by means of a cognition technology. I speculate that the model might help explain several stylized facts that are at odds with an infinite capacity (fully rational) model.

I thank seminar participants at The University of Memphis and specially Bill Smith for helpful comments.
Citation: Nocetti, Diego, (2005) "A Model of Mental Effort and Endogenous Estimation Risk." *Economics Bulletin*, Vol. 4, No. 14 pp. 1−10
Submitted: September 22, 2005. **Accepted:** December 1, 2005.
URL: http://www.economicsbulletin.com/2005/volume4/EB−05D80044A.pdf

1. Introduction

In recent years there has been a flurry of research that departs from the rational agent paradigm in order to explain economic phenomena that are at odds with the standard model. This note tackles one particular departure from the assumption of an omniscient rational agent; namely, the limited use of available information due to scarce cognitive resources and memory deficits.

I take as a starting point the ideas and experimental work of Kahneman (1973) on attention limitations and extend his framework to account for the interaction between attention and memory deficits. In particular, the behavioral model is based on the assumption that individuals optimally allocate their divisible, but limited, attention to estimate parameters of an economic model, by retrieving observations from a stock of memories, by means of a cognition technology.

In the next section I present a general model of attention allocation while, in the framework of an inference problem, section 3 formalizes the interaction between the attention allocation policy and memory limitations. In section 4 I briefly consider some applications of the model. Section 5 compares and contrasts the model with the current literature and concludes.

2. The Allocation of Mental Effort

This section develops a formal model of attention based on the ideas and experimental work of Kahneman (1973). The model is based on the following premises. First, attention (mental effort) is a scarce resource (input). Second, the input is divisible (i.e. processing is parallel as opposed to serial) among activities which might differ in their demands. Third, the effort exerted to a given activity determines a particular output. The "production" of such output is achieved with a given cognition technology. Finally, the allocation of the input is done in an optimal way. Figure 1, which is an adaptation of Kahneman (1973 pg 10), illustrates the interaction between these elements.

To formalize these ideas consider the following framework. At a given instant in time there are *m* actions that demand attention. Let a_i denote the *performance* (consequence) in the *i*-th action for a given level of effort and suppose that the satisfaction, $\mu_i(a_i)$, received from these is separable

$$U = \mu_1(a_1) + \mu_2(a_2) + ... + \mu_m(a_m) \tag{1}$$

where $\mu_i(.)$ is a strictly concave function. Further, suppose that the cognition technology-output relationship is described by

$$f_i(e_i) = a_i \tag{2}$$

where $f_i(e_i)$ represents a (concave) *cognition technology* for a given level of effort. This is subject to the capacity (*k*) constraint

$$e_1 + e_2 + ... + e_m = k \tag{3}$$

Using (2) and (3) we can write the technology constraint as a *cognition possibilities frontier*

$$f_1^{-1}(a_1) + f_2^{-1}(a_2) + ... + f_m^{-1}(a_m) = k \ . \tag{4}$$

The optimal allocation of mental effort implies the following condition,

$$\mu_i^{'}(a_i) - \lambda f_i^{-1\,'}(a_i) = 0 \ \forall i \tag{5}$$

where λ is the Lagrange multiplier and represents the change in the satisfaction received in equilibrium given a small change in the attention capacity constraint. Equation (5) holds for all actions that demand attention; therefore, we have,

$$\frac{f_j^{-1\,'}(a_j)}{f_i^{-1\,'}(a_i)} = \frac{\mu_j^{'}(a_j)}{\mu_i^{'}(a_i)} \ \forall j \ , \tag{6}$$

which is the usual condition that the marginal rate of transformation equals the marginal rate of substitution among all possible actions. Figure 2 illustrates the cognition problem for the case of two actions. The figure shows the optimal tangency condition between the cognition possibilities frontier and the indifference curve and the shift in the frontier (and the optimal allocation of attention) given an increase in capacity from k to k'.

A mundane example might help to grasp the idea of the model. Consider the problem of how to allocate attentional resources to driving down a highway and listening to the radio news. The quality of driving (e.g. correctly change gears) and the amount and quality of the information received from the news broadcast represent the performance on each action. Given a processing constraint, devoting more attention to one action necessarily has a cost in terms of inattention to the other action. For example, paying more attention to driving implies missing part of the news story. On the other hand, devoting more cognitive resources to the story might increase the chances of having an accident.

Now suppose that it starts to rain or that the highway becomes more congested. Because the relative cost of attention to driving has fallen, an increase in the amount of attention to this action is called for. But, how to select how much more attention driving requires? The model proposes that at the optimum we equalize the ratio of marginal utilities (the change in the satisfaction received from a small change in the performance of each action) to the marginal rate of transformation (the performance in one action that the individual gives up to produce a higher performance in the other action). Thus, given identical preferences, a beginner driver will need to reallocate a larger amount of attention than an experienced driver (who has higher productivity).

3. Divided Attention, Memory, and Least Squares

Most economic decisions require estimating parameters that make up an economic model. Recognizing this, a large amount of literature treats individuals as econometricians who base their decisions on a given sample of observations. In this section I analyze an example of the previous framework in which the actions that demand attention take the form of inference problems. Importantly, individuals do not use databases as econometricians do but rely on their memory to infer the parameter estimates. Beyond providing a more realistic flavor to the inference problem, the advantage of such treatment is twofold. First, since the sample size is possibly small due to scarce cognitive resources, parameter uncertainty remains significant even if the data *available* is large and there are no structural shifts. Second, the endogenous characteristic of the sample allows quantifying the magnitude and disentangling the determinants of the deviations from the standard (infinite capacity) model.

The individual has to decide the optimal allocation of attention to make inferences about *m* (uncorrelated) variables that follow the simple process

$$x_{it} = \mu_i + \varepsilon_{it} \tag{7}$$

with ε_{it} is $N \sim \left(0, \sigma_i^2\right)$ and σ_i^2 is known. In the appendix I analyze the case with correlated variables and I show that the main qualitative results that follow remain intact.

Suppose that the individual's objective is to select the estimates, $\hat{\mu}_i$, that minimize the mean square errors. Further, suppose that past observations of each variable compose a stock of memory and that the individual focuses his attention to finding the optimal estimates by retrieving a sample of size n_i from memory. A higher level of effort leads to a larger number of observations and a better performance in the inference problem.

Consider the case where the cognition technology is Cobb Douglas

$$\Phi_i e_i^{\alpha} = n_i \, , \tag{8}$$

where Φ_i is a productivity parameter and $\alpha \leq 1$. In this case the cognition possibilities frontier is a *memory frontier*

$$\left(\frac{n_1}{\Phi_1}\right)^{\frac{1}{\alpha}} + \left(\frac{n_2}{\Phi_2}\right)^{\frac{1}{\alpha}} + ... + \left(\frac{n_m}{\Phi_m}\right)^{\frac{1}{\alpha}} = k \, . \tag{9}$$

It seems reasonable that, for a given level of effort, familiarity increases the ease with which information is retrieved from memory. That is, individuals are relatively more productive retrieving familiar information. Therefore, Φ_i can be interpreted as a familiarity parameter. According to this an individual will, for example, find more difficult to retrieve a list of names randomly selected from the phone book than a list of famous people. Similarly, practice should improve recall. A large amount of literature supports this hypothesis [e.g. Mandler (1980), Gillund and Shiffrin (1984)]

Since, by assumption, the samples are independent the objective function takes the form

$$MAX \ -\sum_{i=1}^{m} \sigma_i^2 / n_i \, . \tag{10}$$

It follows that the optimality condition for the cognition problem is

$$\frac{\sigma_i^2}{n_i^2} = \lambda \frac{1}{\alpha} \left(\frac{n_i}{\Phi_i}\right)^{\frac{1-\alpha}{\alpha}} \frac{1}{\Phi_i} \ \ \forall i \, . \tag{11}$$

This equation says that, at the optimum, the marginal benefit of attention to variable *i* must equal its marginal cost. Alternatively, since this holds for all samples, dividing the marginal benefits of recalling an observation from samples *i* and *j* and the marginal costs we obtain the condition where the marginal rate of substitution equals the marginal rate of transformation.

The optimal sample size is

$$n_i^* = \frac{k^\alpha}{\frac{1}{\Phi_i}\left[1+\left(\sum_{i=1}^{m}\left(\sigma_j^2\Phi_j\right)^{1/(1+\alpha)}\right)\Big/\left(\sigma_i^2\Phi_i\right)^{1/(1+\alpha)}\right]^\alpha} \quad \forall\, j \neq i \tag{12}$$

while the optimal level of effort follows from the production technology and the estimate that minimizes the mean square deviations is $\hat{\mu}_i = \overline{x}_i \sim N\left(\mu_i, \sigma_i^2/n_i^*\right)$.

Corollary. *Attention to variable j, its optimal sample size, and therefore the efficiency of the estimate,*

1. *Increase with capacity*
2. *Increase with its variance and decrease with the variance of all other variables.*
3. *Increase with the productivity of retrieval of this variable and decrease with the productivity of the other variables.*

The intuition for the first two results is straightforward. As capacity increases the individual is able to retrieve more information from memory to solve the inference problem. If capacity were infinite the individual would be able to remember all observations in which case the estimate would equal the actual parameter. Second, an increase in the variance decreases the confidence in the estimate and thus a larger sample is called for. However, because the individuals' attention is constrained this increase will necessarily result in a lower confidence in the other samples.

The fact that an increase in productivity of recall of variable *j* increases the effort invested in this variable and decreases the effort for other variables might seem counterintuitive at first. If, as before, productivity is interpreted as a familiarity parameter this implies that individuals will tend to focus their attention in the retrieval of observations of those variables that are relatively more familiar to them. This is quite surprising since one would expect that, as the familiarity of one variable (action) increases, the individual will distribute his limited capacity in a way that reduces the uncertainty on both variables.

To understand the intuition of the result it is useful to think about the cognition-inference problem with two variables, say A and B. An increase in the productivity of retrieval of observations of variable A acts as an *effective* increase in the resources allocated to this variable. As illustrated in figure 3 this causes a biased expansion of the cognition possibilities frontier in the direction of variable A. Thus, we encounter a type of *Rybczynski effect*, whereby at constant commodity prices (the relative variance of the shocks) an increase in the supply of one input leads to an absolute increase in the output of the good (observations retrieved) that uses this input intensively and an absolute decrease in the output of the other good.

Although a complete treatment of a general equilibrium economy with information sharing is outside the scope of this note a direct extension of this result is that *if* *individuals are allowed to share/trade memories (and risk) they will tend to specialize in the retrieval of those memories with which they are more familiar*. In fact, Becker (1985) derives this result in a model of labor supply where he identifies performance for a given level of effort with wage income. He shows that, under the assumption that women are more productive for non-market activities (e.g. child care), they will tend to specialize in those activities, reducing their wages and participation in the labor market.

Finally, one can also interpret these results in terms of any actions that demand attention such as my previous example of driving and paying attention to the radio news. In some situations, however, the result of a higher (lower) level of attention for more (less) familiar actions is counterfactual. For example, in identical environments experienced drivers are generally able to perform other actions better than novice drivers. This situation, however, is easily accommodated by the model if increases in performance have a limit. In such situation, after performance in one action has reached the limit any further increase in productivity will allow the individual to allocate more effort to other concurrent activities.

4. Applications

In this section I briefly speculate on the applicability of the model. The applications are based on research that is under way (Nocetti, 2005) and I wish that many more will follow.

4.1 Portfolio Choice

There is a large literature, mostly in a Bayesian framework, which considers portfolio choice under parameter uncertainty [Zellner and Chetty (1965) is the seminal paper on the subject]. As one would expect, estimation risk reduces the optimal equity share. The endogeneity of estimation risk in the present framework might provide additional interesting insights. For example, more familiar equities (with higher productivity of recall) will have lower estimation risk and therefore the holdings on those equities will be larger. This might help explain the well known *home bias puzzle* as well as the lack of intra-national diversification.

4.2 Asset Pricing

Timmerman (1993) shows that stock returns display excess volatility if individuals have to estimate the growth rate of dividends. Intuitively, the dividend yield is not constant as in models with “rational” agents (infinite capacity in the present framework).

The present model predicts, in addition, that small and lesser known stocks are more volatile as observed in the data [e.g. Pastor and Varonesi (2003) and Brown and Ferreira (2004)]. Intuitively, if the representative investor is more productive recalling more familiar information the volatility of the estimates and therefore the volatility of the dividend yield will be smaller.

Further, although not analyzed here, any predictable bias of memory retrieval introduces stock returns predictability. For example, it is well established in the psychology literature that individuals tend to recall information that matches in valence the mood at the time of retrieval (known as *mood-congruence effect*). If one can predict moods stock returns will be predictable as well [in periods of optimism (pessimism) investors will overestimate (underestimate) the growth rate of dividends]. It is an empirical question whether stock returns are related to, for example, the index of consumer sentiment which is indeed highly serially correlated.

4.3 Consumption

Similar to the case of asset prices, the model might provide an explanation to the observed failure of the random walk hypothesis of consumption and in particular to Carroll, Fuhrer, and Wilcox (1994) findings that “changes in sentiment not only forecast changes in spending, but also cause them”. In fact, in line with the present model where estimation risk is endogenous, they argue “If consumer sentiment is, in part, a measure of uncertainty, one might hope that a model of precautionary saving would be consistent with our results”.

4.4 Inflation Expectations and Labor Supply

The present model, in which effort is allocated to infer parameter values, might complement Becker’s (1985) model of labor supply. For example, effort exerted to market activities might compete with the demand for attention of inferences about the pattern of economic variables. It is well known, for example, that labor productivity is negatively related with the variability of inflation. The present model suggests a possible explanation. When inflation is more variable people allocate more attention to infer its future pattern and less to market activities.

5. Conclusion

A number of papers have recently explored the implications of limited attention in a variety of settings. For example, Gabaix and Laibson (2004) present a model in which agents optimally allocate thinking time. In their model individuals follow simple heuristics and, based on an exogenous cost of thinking, decide which good deserves the full capacity of attention at a given point in time (i.e. processing is serial). In contrast, the present note stresses that attention is limited but divisible[1]. The allocation of cognitive resources also involves a trade-off among different cognitive operations with a corresponding opportunity cost. However, instead of assuming that cognitive operations involve an exogenous cost, because more than one action can be performed concurrently, the cost is endogenously given by a loss in attentiveness (and therefore performance) to other operations.

The hypothesis presented here is also close to Sims’ (2003) model of *Rational Inattention*. Sims uses the tools of information theory to develop the idea that economic agents decide optimally the amount of noise they receive. Information theory deals with the question of *how much* information can be processed at a given instant in time while Sims explores *what* information is acquired. In addition, I consider *how* the information is processed (a cognition technology). For example, I propose that individuals are more productive processing familiar information and derive important implications from such behavior.

Becker (1985) provided an early treatment of *how* people invest a limited amount of effort into competing activities. Since Kahneman’s (1973) work provides a psychological foundation to Becker’s introspection, the model of attention allocation in the present note

[1] Pashler and Johnston (1998) summarize experimental evidence on attention limitations.

is similar to the latter. Beyond providing further applicability to Becker's model, focusing in the interaction between attention and memory deficits offers a well documented psychological foundation.

Only a few papers have explored the implications of memory deficits for economic behavior. Closest to the present paper, Mullainathan (2002) presents a model in which agents make inferences about economic variables by applying Bayes' rule to the recalled history as if it were the true history. Mullainathan focuses the analysis on a number of well known biases and assumes that individuals follow mechanical rules. Instead, this note stresses that memory retrieval is a byproduct of an optimization problem. The advantage of the present context is the characterization of how the cognition problem changes with the environment while a possible drawback is the over-simplification of the cognition procedure.

The behavioral model is simple and there are many possible extensions: considering more sophisticated inference problems, analyzing memory biases, incorporating information sharing, etc. I hope that other researchers will allocate at least part of their attention to extent the present study.

References

Becker, G. (1985) "Human Capital, Effort, and the Sexual Division of Labor" *Journal of Labor Economics*, Vol. 3, No. 1, Part 2: Trends in Women's Work, Education, and Family Building. pp. S33-S58

Brown, David and Ferreira Miguel (2005) "Information in the Idiosyncratic volatility of Small Firms". Mimeo. Brown University.

Carroll, C., Fuhrer, J., and Wilcox, D. (1994) "Does Consumer Sentiment Forecast Household Spending? If so, Why?" *The American Economic Review* 84 1397-1408

Gabaix, X. and Laibson, D. (2004) "Bounded Rationality and Directed Cognition" Mimeo, MIT and Harvard.

Gillund G and Shiffrin R.M. (1984) "A Retrieval Model for Both Recognition and Recall" *Psychological Review*, 91, 1-65

Kahneman, D. (1973) "*Attention and Effort*". Prentice Hall, New Jersey.

Mandler G. (1980) "Recognizing: the Judgment of Previous Occurrence" *Psychological Review*, 87, 252-271

Mullainathan, S. (2002) "A Memory Based Model of Bounded Rationality" *Quarterly Journal of Economics* 117 735-757

Nocetti, D. (2005) "Mental Effort, Memory Deficits, and Financial Markets" Mimeo The University of Memphis.

Pashler, H. and Johnston J. (1998) "Attentional Limitations in Dual-Task Performance,in Pashler, ed., *Attention*, Psychology Press.

Pastor, L., and Veronesi P. (2003) "Stock Valuation and Learning About Profitability" *Journal of Finance* 58, 1749-1789.

Sims, C. (2003) "Implications of Rational Inattention" *Journal of Monetary Economics*, 50(3), 665-690.

Zellner, Arnold and Chetty, Karuppan (1965) "Prediction and Decision Problems in Regression Models from the Bayesian Point of view" *Journal of the American Statistical Association* 60, 608-616

APPENDIX

Optimal Allocation of Attention with Correlated Variables

Suppose that, as before, there are *m* variables that demand attention and follow the process

$x_{it} = \mu_i + \varepsilon_{it}$.

Now, however, suppose that the shocks to variable *i* are correlated with the shocks to variable *j* and denote their correlation coefficient r_{ij} . Then, minimizing the mean square errors is equivalent to

$$MAX \ -\sum_{i=1}^{m}\sum_{z=1}^{m}\frac{r_{iz}\sigma_i\sigma_z}{n_i} .$$

Suppose that the cognition technology is the same as before. Then, while the optimal estimate is the sample mean of each variable, the optimal sample size, and thus the variance of the estimate, will differ. In particular, the optimality condition for the cognition problem is

$$\frac{\sum_{z=1}^{m} r_{iz}\sigma_i\sigma_z}{n_i^2} = \lambda\frac{1}{\alpha}\left(\frac{n_i}{\Phi_i}\right)^{\frac{1-\alpha}{\alpha}}\frac{1}{\Phi_i} \ \ \forall i ,$$

from which it follows that the optimal sample size is

$$n_j^* = \frac{k^\alpha}{\frac{1}{\Phi_j}\left[1+\left(\sum_{i=1}^{m}\left(\Phi_i\sum_{z=1}^{m} r_{iz}\sigma_i\sigma_z\right)^{1/(1+\alpha)}\right)\Big/\left(\Phi_j\sum_{z=1}^{m} r_{jz}\sigma_j\sigma_z\right)^{1/(1+\alpha)}\right]^\alpha} \quad \forall\ i \neq j$$

The comparative statics are easier to interpret in the case of two variables, say *i* and *j*, in which case the optimal sample size reduces to

$$n_j^* = \frac{k^\alpha}{\frac{1}{\Phi_j}\left[1+\left(\frac{\Phi_i}{\Phi_j}\frac{\sigma_i^2+\sigma_{ij}}{\sigma_j^2+\sigma_{ji}}\right)^{1/(1+\alpha)}\right]^\alpha} .$$

An increase in the covariance, σ_{ij} , increases (decreases) the optimal sample size of variable *j* (*i*) IFF the variance of variable *i* is larger that the variance of variable *j* (i.e. $\sigma_i^2 > \sigma_j^2$), provided $\sigma_{ij} > 0$, and it will have the opposite effect (given an increase in the absolute value of σ_{ij}) if $\sigma_{ij} < 0$. The intuition follows directly from the case with uncorrelated variables. For example, if $\sigma_{ij} > 0$ and increases the individual reallocates her attention towards the now relatively more volatile variable and away from the other variables.

Figure 1. Process of Allocation of Mental Effort

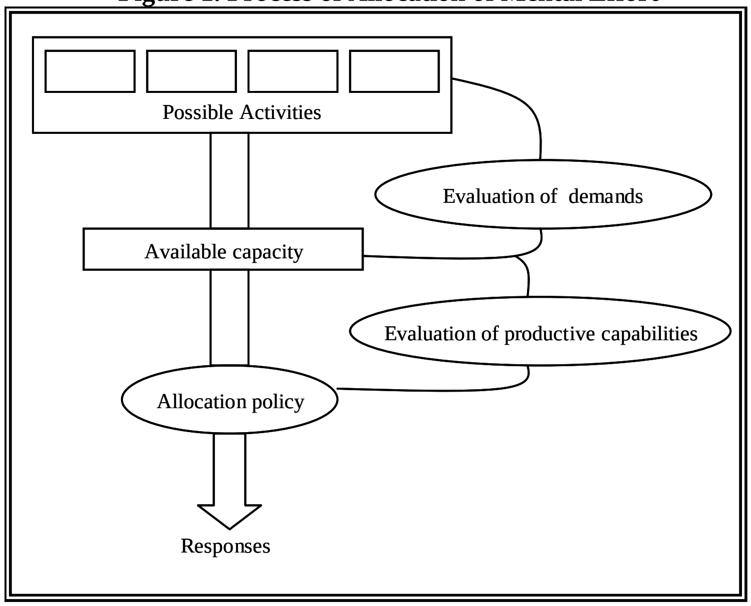

Figure 2. Optimal Allocation of Attention

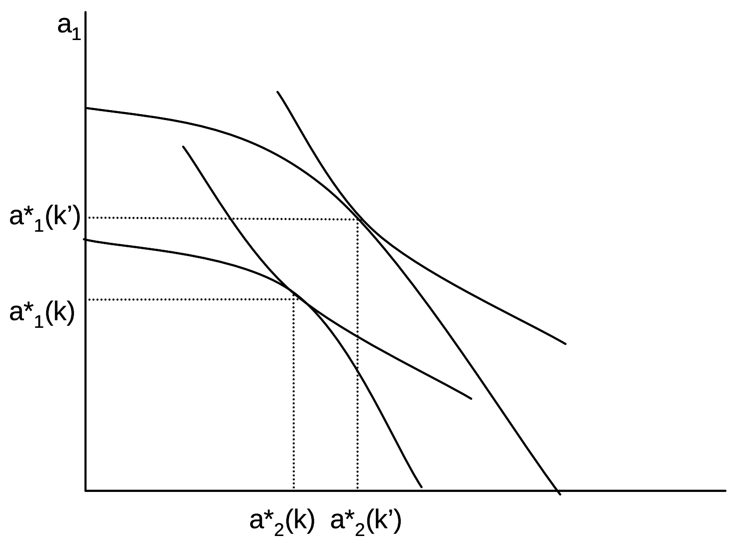

Figure 3. Increase in Productivity of Memories Retrieval of one Variable