The Devil You Know: Rational Inattention to Discrete Choices when Prior Information Matters

Bruno Pellegrino

Columbia Business School, CESifo

July 16, 2023

Abstract

In the seminal rational inattention model of Matějka and McKay (2015), logit demand arises from the discrete choice of agents who are uncertain about choice payoffs and who have access to a flexible, costly information acquisition technology (RI-logit). A notable limitation of this powerful framework is the lack of known general closed-form solutions, allowing the decision maker's prior information to differ across choices. In this paper, I solve the RI-logit model analytically for a large family of priors known as multivariate *Tempered Stable* (TS) distributions. In my analytical formulation, decision makers can be biased, display aversion to prior uncertainty, and thus tend to select choices that are familiar (i.e. for which they hold a less disperse prior). This result allows to study how the RI-logit choice probabilities react to an exogenous change in prior information, thus extending the model's applicability to a new range of settings where prior information matters. I provide one such application: I show how to use the closed-form RI-logit model to study the behavior of risk-averse investors who select risky projects in an environment characterized by epistemic uncertainty (risk-adjusted expected returns are unknown, but can be learnt at a cost).

Keywords: Rational Inattention, Discrete Choice, Logit, Information Acquisition, Uncertainty.
JEL Codes: D11, D81, D83

bruno.pellegrino@icloud.com | www.brunopellegrino.com
University of Maryland, Robert H. Smith School of Business, College Park, MD 20742.
I thank Filip Matejka and Alisdair McKay for valuable feedback.

1 Introduction

How do rational agents make decisions under uncertainty when information is costly? What type and amount of information will they process before making a decision? This is the research question that lies at the heart of the economics literature on rational inattention (Sims, 2003).

Within this literature, the rational inattention multinomial logit (RI-logit) model has established itself as the benchmark tool for modeling discrete choice with costly information acquisition. Introduced by Matějka and McKay (2015), this framework characterizes the behavior of individuals who choose an option (among many) in an environment characterized by uncertain payoffs, and who are able to (partly) resolve this uncertainty by acquiring costly information.

The appeal of this framework lies in its combination of generality, elegance and tractability. In this model, agents are endowed with prior beliefs over the distribution of payoffs, and are able to acquire additional information about the payoffs in the form of a signal. The information acquisition problem is "unrestricted", in the sense that the agents are not limited to choosing signals with a pre-determined probability distribution. The agents are effectively able to choose a distribution for the signal conditional on the state vector of payoffs. In other words, the decision maker has agency not only over the precision of the signal, but over the entire information structure. The cost of the signal is proportional to the informativeness of the signal, which is quantified using the canonical Entropy formula due to Shannon (1948).

The framework yields state-conditional choice probabilities that take the familiar multinomial logit form (McFadden, 1973). One of the key differences between MM's micro-foundation and McFadden's is that MM's logit formula incorporates information from the agents' prior. In other words, choices that appear attractive ex-ante (i.e. before the agents acquire any signals) are more likely to be selected, irrespective of the true underlying state vector of payoffs. This makes MM's model in some sense more general. When the choices are ex-ante undistinguishable, the model simplifies to the canonical multinomial logit.

Despite its theoretical appeal, the model faces an important constraint in its applicability, namely the lack of known closed-form solutions for the general case, where agents may possess better prior information about particular choices. To be more specific, in MM the *conditional* choice probabilities (which depend on the true payoffs) are written as a function of the corresponding *unconditional* choice probabilities; those in turn depend on the particular probability distribution that is assumed to describe the agent's prior beliefs.

At present, no distributional assumption on the prior is known to produce closed-form solutions for these probabilities, at least for the case where the prior is allowed to be asymmetrically-precise. If analytical solutions can be derived, they would broaden the applicability of the MM framework to situations where choices are not equal ex-ante, and the agents' prior knowledge may thus play a role. Additionally, these solutions would enable closed-form comparative statics on the prior. An intriguing question to explore is, for example, how do choice probabilities tilt when the decision-maker receives better prior information about a specific option? Some notable settings where the agent's prior information is likely to matter include healthcare insurance choice (Brown and Jeon, 2020) and international investment (Pellegrino, Spolaore, and Wacziarg, 2021).

This paper provides an analytical solution to the RI-Logit model that obtains when the agent's prior belongs to a large family of multivariate probability models known as multivariate Tempered Stable (TS) distributions. This family includes several well-known probability models such as the Gamma and the Inverse Gaussian distributions, along with their special cases (Exponential, Chi-squared, Erlang...). Under this assumption, we obtain closed-form solutions for the unconditional and conditional probabilities. The precision of the prior distribution is allowed to differ across choices, and prior uncertainty is reflected in the conditional choice probabilities. In addition, my closed-form solution allows agents to be biased towards specific choices.

The derived closed-form solution suggests that, in accordance with the familiar adage, optimizing agents prefer "the devil they know" over that which they don't know. In other words, everything else being equal, agents are more likely to choose options they have better prior knowledge about, as captured by the accuracy of their prior information; this behavior reflects an aversion to prior uncertainty. After presenting the solution, I offer an explanation for why agents exhibit aversion to prior uncertainty and discuss how it affects the ability to infer preference orderings from choice data.

My closed-form formulation of the RI-logit model allows for a more realistic representation of decision-making in environments where agents are not equally informed about all choices, thus expanding the range of applications of RI-logit.

After deriving and discussing this solution, I present a practical application. One of the lesser-known advantages of the RI-logit model is that it can be used to model risk as well as epistemic uncertainty. Risk (or aleatoric uncertainty) is uncertainty about the outcome of a random process. Epistemic uncertainty is the lack of knowledge about the true values of the parameters of a stochastic process, which could potentially be learnt. I present a model where risk-averse agents allocate capital among several firms that are subject to unlearnable productivity shocks. In addition to the risk presented by these shocks, the agents face epistemic uncertainty about (and are rationally-inattentive to) the risk-return properties of the assets.

I show how capital allocation and the assets' expected returns are determined in equilibrium, and how these are affected by (prior) epistemic uncertainty. Three salient insights of this simple model are: 1) assets pay a return premium for epistemic uncertainty; 2) epistemic uncertainty makes asset demand inelastic; 3) epistemic uncertainty leads to a real misallocation of capital that can be quantified.

This paper contributes to a growing body of work, originally pioneered by Sims (2003), that develops and utilizes models of decision making under rational inattention. Maćkowiak, Matějka, and Wiederholt (2021) provide a review of this literature. In related work, Dasgupta and Mondria (2018) provide a closed-form solution to the RI-logit model that allows for homogeneous precision and heterogeneous mean in the prior; however (as explained by the authors) this solution requires the highly-stringent assumption that the parameters controlling the shape of the prior and the information cost must be the same.

The rest of the paper is structured as follows. In the second section, I briefly review the Rational Inattention Logit model. In the third section, I present my closed-form solution to the model and show how the resulting choice probabilities satisfy the necessary and sufficient conditions for a solution. In the fourth section, I present the application to capital allocation under epistemic uncertainty. In the fourth section, I conclude.

2 The Rational Inattention Multinomial Logit Model

In this section, I review the rational inattention multinomial logit (RI-logit) model of Matějka and McKay (2015, henceforth MM).

Let there be an agent who must choose a single action $i \in \{1, 2, \ldots, n\}$. The environment is characterized by uncertainty, which we model using a state vector $\boldsymbol{\omega} \in \mathbb{R}^n$. The payoff to the agent from choosing action i under state $\boldsymbol{\omega}$ is the i'th entry of the state vector, ω_i. The decision maker's objective is to select the action that yields the highest payoff in expectation.

To streamline notation, I use uppercase letters to denote the exponential transform of a variable denoted by the corresponding lowercase - for instance:

$$\Omega_i \stackrel{\text{def}}{=} \exp(\omega_i) \tag{2.1}$$

Note that this setup can be used to model risk as well as epistemic uncertainty. The only difference between these two framings of the model lies in whether the state vector $\boldsymbol{\omega}$ obtains before or after the agent makes their decision. In both cases, the decision maker is uninformed about the state of nature, and thus we will ignore this difference for now. We will come back to it in Section 4.

The decision maker is Bayesian: he or she is endowed with prior beliefs about the state vector $\boldsymbol{\omega}$. These prior beliefs are described by a probability distribution $G \in \Delta(\mathbb{R}^n)$, where $\Delta(\mathbb{R}^n)$ is the set of all probability distributions over $\mathbb{R}^n$. The problem they solve is comprised of two stages. In the first stage, the decision maker has access to an information acquisition technology that enables them to obtain a noisy signal $\mathbf{s} \in \mathbb{R}^n$: the agent chooses the state-conditional distribution of this signal. Because the signal is drawn from a distribution that conditions on the actual state $\boldsymbol{\omega}$, the signal is (potentially) informative about the state vector.

In the second stage, after observing the signal, the agent updates their prior into a posterior distribution $F \in \Delta(\mathbb{R}^n)$ and chooses an action $i \in \{1, \dots, n\}$, with the objective of maximizing the expected payoff:

$$V(F) \;=\; \max_{i \in \{1,\dots,n\}} \mathbb{E}^F(\omega_i) \tag{2.2}$$

where the superscript $\left(^F\right)$ denotes that the expectation is taken under the posterior probability measure $(\mathrm{d}F)$. This payoff is a function of the posterior distribution F. Thus, formally $V : \Delta(\mathbb{R}^n) \to \mathbb{R}$.

We now outline the information acquisition problem, which the decision maker solves in the first stage. Importantly, the decision-maker has agency not only on the "amount" of information, but also on the information structure. At this stage, the decision maker's task is to select a joint distribution for the signal and the state $F(\mathbf{s}, \boldsymbol{\omega}) \in \Delta(\mathbb{R}^n \times \mathbb{R}^n)$, subject to the restriction that the marginal distribution over states equals the decision maker's prior G (i.e. the agent does Bayesian updating). The choice of this joint distribution induces, upon the realization of the signal, the formation of a posterior, which I denote by $F(\boldsymbol{\omega}|\mathbf{s})$.

The signal is not free. The agent pays a utility cost $C(F(\boldsymbol{\omega}|\mathbf{s}), G(\boldsymbol{\omega}))$ for acquiring the signal, which depends on the prior and the posterior probability distributions of the state.

Hence, the mathematical formulation of the first-stage maximization problem is:

$$\max_{F \in \Delta(\mathbb{R}^n \times \mathbb{R}^n)} \int_{\boldsymbol{\omega} \in \mathbb{R}^n} \int_{\mathbf{s} \in \mathbb{R}^n} V(F(\cdot|\mathbf{s}))\; \mathrm{d}F(\mathbf{s}|\boldsymbol{\omega})\; \mathrm{d}G(\boldsymbol{\omega}) - C(F(\boldsymbol{\omega}|\mathbf{s}), G(\boldsymbol{\omega})) \tag{2.3}$$

$$\text{subject to :} \quad \int_{\mathbf{s} \in \mathbb{R}^n} f(\mathbf{s}, \boldsymbol{\omega})\, \mathrm{d}\mathbf{s} \;=\; G(\boldsymbol{\omega}) \quad \forall \boldsymbol{\omega} \in \mathbb{R}^N \tag{2.4}$$

The constraint in (2.4) simply says that F and G satisfy Bayes's rule.

As standard in this literature since Sims (2003), I assume that the cost function $C(\cdot)$ is proportional to the information content of the signal, measured as the expected reduction in Shannon Entropy (H) from Bayesian updating:

$$C(F(\boldsymbol{\omega}|\mathbf{s})) \;\overset{\text{def}}{=}\; \lambda \cdot \{\mathrm{H}(G(\boldsymbol{\omega})) - \mathbb{E}_{\mathbf{s}}[\mathrm{H}(F(\boldsymbol{\omega}|\mathbf{s}))]\} \tag{2.5}$$

The parameter λ controls the cost of acquiring information. In what follows, to keep the problem non-trivial, I shall assume that $\lambda > 0$.

We define $\mathcal{P}_i(\boldsymbol{\omega})$, the *conditional probability* that the agent chooses action i, where the conditioning is on the state of nature $\boldsymbol{\omega}$:

$$\mathcal{P}_i(\boldsymbol{\omega}) \;\overset{\text{def}}{=}\; \int_{\mathbf{s} \in \mathcal{S}(i)} \mathrm{d}F(\mathbf{s}|\boldsymbol{\omega}) \tag{2.6}$$

where $\mathcal{S}_i$ is the set of realizations of the signal $\mathbf{s}$ that cause the agent to select action i. Let us also define $\mathcal{P}_i^0$, the *unconditional probability* that the decision maker selects action i. By definition, it is obtained by integrating the conditional probability over the prior G:

$$\mathcal{P}_i^0 \;\overset{\text{def}}{=}\; \int_{\boldsymbol{\omega} \in \mathbb{R}^n} \mathcal{P}_i(\boldsymbol{\omega})\, \mathrm{d}G(\boldsymbol{\omega}) \tag{2.7}$$

By Theorem 1 in MM, the agent's optimization of (2.3) implies that the state-conditional probability of choosing action i takes the following logit-like form:

$$\mathcal{P}_i(\boldsymbol{\omega}) \;=\; \frac{\Omega_i^{\frac{1}{\lambda}}\ \mathcal{P}_i^0}{\sum_{j=1}^{n} \Omega_j^{\frac{1}{\lambda}}\ \mathcal{P}_j^0} \qquad \text{(almost surely)} \tag{2.8}$$

The equation above displays two key features. First, decision makers are more likely to select more valuable choices, but they are not guaranteed to do so: the elasticity of $\mathcal{P}_i(\boldsymbol{\omega})$ with respect to ω_i is not infinite, and is inversely proportional to the cost of information λ. In other words, the higher this cost, the less information agents acquire, and the less likely it is that they select the best options. Second, the decision maker's prior information affects their choice probabilities through the unconditional probability $\mathcal{P}_i^0$, which in turn depends on the prior beliefs G. This implies that choices that appear ex-ante more attractive are more likely to be selected, irrespective of their fundamental payoff ω_i. Note, however, that we do not know at this stage what "ex-ante more attractive" means, as we have not yet specified a functional form for G.

Next, to simplify notation, I follow Steiner, Stewart, and Matějka (2017) and normalize $\lambda = 1$. This equates to operating a change of variable $\boldsymbol{\omega}' = \frac{1}{\lambda}\boldsymbol{\omega}$ and omitting the ($'$) symbol. This change of variable does not affect the information cost because, by the property of the entropy operator (for a generic probability distribution Q) $\mathrm{H}\left(Q\left(\boldsymbol{\omega}\right)\right) = \mathrm{H}\left(Q\left(\boldsymbol{\omega}\right)\right) + \log\lambda$.

As a result of this normalization. we can re-write equation (2.8) as:

$$\mathcal{P}_i(\boldsymbol{\omega}) \;=\; \frac{\Omega_i\ \mathcal{P}_i^0}{\sum_{j=1}^{n} \Omega_j\ \mathcal{P}_j^0} \tag{2.9}$$

While condition (2.9) is necessary for a solution to Problem (2.3)-(2.4), it is not sufficient. MM add the condition that the unconditional probabilities solve:

$$\begin{aligned} &\max_{\mathcal{P}^0} \int_{\boldsymbol{\omega}\in\mathbb{R}^n} \log\left(\sum_{j=1}^{n} \Omega_j \cdot \mathcal{P}_j^0\right) \mathrm{d}G\left(\boldsymbol{\omega}\right) \\ \text{s.t.} \qquad &\mathcal{P}_j^0 \in [0,1] \;\text{ for } i = 1,2,...,n \qquad \text{and} \qquad \sum_{j=1}^{n} \mathcal{P}_j^0 = 1 \end{aligned} \tag{2.10}$$

conditions in (2.9) and (2.10) are jointly necessary and sufficient for a solution to the problem in (2.3). In their Online Appendix C, MM show that, when all options are selected with positive probability ($\mathcal{P}_j^0 > 0$ for $i = 1,2,...,n$) this condition equates to

$$\int_{\boldsymbol{\omega}\in\mathbb{R}^n} \frac{\Omega_i \cdot \mathcal{P}_i^0}{\sum_{j=1}^{n} \Omega_j \cdot \mathcal{P}_j^0} \cdot \mathrm{d}G\left(\boldsymbol{\omega}\right) \;=\; \mathcal{P}_j^0 \qquad \text{for all } i \tag{2.11}$$

Despite their theoretical importance, the above results do not lead directly to an analytical characterization of the choice probabilities. In order to obtain one, we must solve for $\mathcal{P}_i^0$ first. This in turn requires making an assumption about the the prior beliefs G.

MM show that, if the decision maker's prior beliefs are exchangeable in i (that is, if choices are ex-ante identical), the unconditional probabilities $\mathcal{P}_i^0$ simplify out from the expression above: this yields the classical multinomial logit model. Unlike McFadden (1973), who derived it from extreme value theory, MM obtain the multinomial logit model from an information theory micro-foundation. I refer to this solution as the "trivial" solution, in the mathematical sense that the agents' prior information does not have any impact on the choice probabilities.

Dasgupta and Mondria (2018) provide a closed-form solution for a limit case where the priors follow a one-parameter Cardell C-distribution, although (as discussed by the authors) this formulation requires the knife-edge restriction that the C-distribution's only parameter coincides with the cost of information (λ). Furthermore, this configuration does not permit heterogeneity in the prior dispersion across different choices.

3 Prior Uncertainty and Bias: Analytical Results

The current lack of closed-form solutions for the case of asymmetrically-precise priors implies that we can't exogenously augment or diminish the "amount" of choice-specific information that is available to the decision maker (at least, not without resorting to numerical methods). In other words, we are limited in our ability to study how a decision maker's choice probabilities tilt when we supply them with "better" or "worse" information about some specific choice i. One can easily opine that one of the most attractive properties of RI-logit is that it allows information to enter the model endogenously as well as exogenously (through the prior). Thus, it is easy to see how a closed-form solution that allows for flexibility in the prior would vastly extend the applicability of MM's framework.

My objective in this section is to derive a closed-form solution that allows for: 1) asymmetric prior precision: that is, I want to allow the agent's prior distribution to be unequally dispersed across choices; 2) bias: that is, I want to allow the agents to be systematically over-optimistic about specific choices.

3.1 Incorporating Bias

I start by slightly generalizing the model of Section 2 to allow bias to creep in the decision maker's information choice problem. Specifically, I assume that the decision maker is no longer able to condition the signal directly on the payoff vector $\boldsymbol{\omega}$. Instead, they can condition the signal on a vector $\tilde{\boldsymbol{\omega}} \in \mathbb{R}^n$, which is obtained by perturbing the payoff vector with an additive shock $\boldsymbol{\delta} \in \mathbb{R}^n$:

$$\tilde{\omega}_i \overset{\text{def}}{=} \omega_i - \delta_i \tag{3.1}$$

The shifter δ_i represents the agent's degree of over-optimism with respect to choice i. The decision maker thus has prior beliefs $G(\tilde{\boldsymbol{\omega}})$ and selects a joint distribution $F(\tilde{\boldsymbol{\omega}}, \mathbf{s})$. As a consequence of this modification, the set of state variables now also includes $\boldsymbol{\delta}$ and the state space is $\mathbb{R}^{2n}$. That is, the combined state is $(\boldsymbol{\omega}, \boldsymbol{\delta}) \in \mathbb{R}^{2n}$.

I assume that the shifter itself is unlearnable by the agents (i.e. they can't condition the signal on $\boldsymbol{\delta}$, as it would be akin to being able to condition on $\boldsymbol{\omega}$). The agent's belief is that δ_i has identical mean (which we impose equal to zero without loss of generality) and that it follows a probability distribution that is independent of the payoff vector $\boldsymbol{\omega}$:

$$\mathbb{E}(\delta_i|\boldsymbol{\omega}) = 0 \qquad \forall\, i = 1, 2, ..., n \tag{3.2}$$

Then, the second stage problem in (2.2) can be re-written, using the law of iterated expectations, as:

$$V(F) = \max_{i\in\{1,\dots,n\}} \mathbb{E}^F(\omega_i) = \max_{i\in\{1,\dots,n\}} \mathbb{E}^F[\mathbb{E}(\omega_i|\tilde{\omega}_i)] = \max_{i\in\{1,\dots,n\}} \mathbb{E}^F(\tilde{\omega}_i) \tag{3.3}$$

In other words, assumption (3.2) implies that the decision maker solves the exact same first-stage and second-stage optimization problems as in (2.2) and (2.3): the only difference is that she targets the perturbed payoffs ($\tilde{\boldsymbol{\omega}}$), instead of $\boldsymbol{\omega}$. We can simply replace $\boldsymbol{\omega}$ with $\tilde{\boldsymbol{\omega}}$ in the information acquisition cost in (2.5), as a consequence of the fact that the Shannon Entropy operator H is unaffected by mean shifts in the distribution of $\boldsymbol{\omega}$. If $\delta_i = 0$ for all i, the decision maker's problem collapses back to that described in Section 2.

Then, the conditional choice probabilities satisfy the following variant of equation (2.9)

$$\mathcal{P}_i(\boldsymbol{\omega},\boldsymbol{\delta}) \;=\; \frac{\tilde{\Omega}_i\ \mathcal{P}_i^0}{\sum_{j=1}^n \tilde{\Omega}_j\ \mathcal{P}_j^0} \;\equiv\; \frac{\Omega_i\ \Delta_i\ \mathcal{P}_i^0}{\sum_{j=1}^n \Omega_j\ \Delta_j\ \mathcal{P}_j^0} \tag{3.4}$$

the unconditional probability $\mathcal{P}_i^0$ is now defined as:

$$\mathcal{P}_i^0 \stackrel{\text{def}}{=} \int_{\tilde{\boldsymbol{\omega}}\in\mathbb{R}^n} \mathcal{P}_i(\tilde{\boldsymbol{\omega}})\ \mathrm{d}G\left(\tilde{\boldsymbol{\omega}}\right) \tag{3.5}$$

and condition (2.10) becomes

$$\begin{aligned} &\max_{\mathcal{P}^0} \textstyle\int_{\boldsymbol{\omega}\in\mathbb{R}^n} \log\left(\sum_{j=1}^n \tilde{\Omega}_j \cdot \mathcal{P}_j^0\right) \mathrm{d}G\left(\boldsymbol{\omega}\right) \\ &\text{s.t.}\qquad \mathcal{P}_j^0 \in [0,1] \ \text{ for } i=1,2,...,n \qquad \text{and} \qquad \textstyle\sum_{j=1}^n \mathcal{P}_j^0 = 1 \end{aligned} \tag{3.6}$$

As can be seen from equation (2.9), the conditional choice probabilities are now affected not only by fundamentals (ω_i) and by the unconditional probabilities $\left(\mathcal{P}_i^0\right)$, but also by bias ($\delta_i$). This behavior is rational, in that the decision maker is unable to learn about the state vector ($\boldsymbol{\omega},\boldsymbol{\delta}$) directly: they can only learn about the vector of perturbed payoffs ($\tilde{\boldsymbol{\omega}}$). The consequence is that the ex-post choice probabilities tilt in favor of choices towards which the agent is biased (higher Δ_i).

3.2 Prior Beliefs and Closed-Form Solution

Next, in order to derive closed-form solutions for $\mathcal{P}_i(\boldsymbol{\omega},\boldsymbol{\delta})$ and $\mathcal{P}_i^0$, I make a functional form assumption about the agent's prior beliefs for $\tilde{\boldsymbol{\omega}}$. I focus on a class of continuous prior densities known as *Tempered Stable Distributions* (Kolossiatis et al., 2011). This is a sub-class of the important family of distributions described by Tweedie (1984).

Tweedie distributions can be defined as exponential dispersion models (Jørgensen, 1987) that have a power variance function. This family of distributions is indexed by their expectation μ, a dispersion parameter σ^2 and a *power index* $p \in (-\infty,0] \cup [1,\infty)$. Their defining restriction of having a "power variance function" simply means that, for a random variable that follows a Tweedie-family distribution with power index p, the variance must equal to $\sigma^2\mu^p$. Formally:

$$X \;\sim\; \mathrm{Tw}_p\left(\mu,\sigma^2\right) \quad\Longrightarrow\quad \mathrm{var}\left(X\right) \;=\; \sigma^2\mu^p \tag{3.7}$$

In what follows, I refer to the reciprocal of σ^2 as the "precision" parameter.

Tweedie-family distributions are scale-invariant, and include several well-known probability models such as the Gaussian ($p=0$) and the Poisson ($p=1$). The subset that is of interest to us, that of Tempered Stable (TS) distributions, is the set of Tweedie-family distributions that have power index larger than two[1]: these are defined over $\mathbb{R}_+$ and have no mass point at zero. This subset includes the Gamma Distribution ($p=2$) and the Inverse Gaussian Distribution ($p=3$) along with their special cases (Exponential, Chi-Squared, Erlang, etc...).

A key contribution of this paper is to show that, when the prior follows a TS distribution with common mean and heterogeneous precision, we obtain an intuitive, non-trivial closed-form solution for $\mathcal{P}_i^0$ and $\mathcal{P}_i(\boldsymbol{\omega},\boldsymbol{\delta})$.

[1]See Jørgensen (1987, section 3), who refers to these distributions as Tweedie distributions "generated by positive stable distributions", and Brix (1999, section 2), who refers to the same distributions as "Generalized Gamma" measures.

Proposition 1. *If the agent's prior beliefs G are such that $\tilde{\Omega}_i$ follows a multivariate probability distribution from the Tweedie family with independent draws over i, uniform power index $p \geq 2$ (tempered stable), uniform mean $\mu > 0$, and choice-specific precision $\mathcal{I}_i^G > 0$:*

$$\tilde{\Omega}_i \overset{G}{\sim} \mathrm{Tw}_p\left(\mu, \frac{1}{\mathcal{I}_i^G}\right) \qquad p \geq 2 \tag{3.8}$$

Then, it follows that:

$$\mathcal{P}_i^0 > 0 \qquad \text{and} \qquad \mathcal{P}_i(\boldsymbol{\omega}, \boldsymbol{\delta}) > 0 \qquad \forall\, i \in \{1, 2, ..., n\} \tag{3.9}$$

the unconditional choice probabilities are given by:

$$\mathcal{P}_i^0 \;=\; \frac{\mathcal{I}_i^G}{\sum_{j=1}^n \mathcal{I}_j^G} \tag{3.10}$$

and the following analytical expression obtains for equation (3.4)*:*

$$\mathcal{P}_i(\boldsymbol{\omega}, \boldsymbol{\delta}) \;=\; \frac{\Omega_i\ \Delta_i\ \mathcal{I}_i^G}{\sum_{j=1}^n\ \Omega_j\ \Delta_j\ \mathcal{I}_j^G} \tag{3.11}$$

Proof. In the Appendix, I prove that (3.10) and (3.11) satisfy conditions (3.4) and (3.6), (which are necessary and sufficient by Lemma 2 of MM) as well as (3.9). □

Notice that this solution applies for any value of the power index p and the mean μ.

In the solution above, the unconditional probability $\mathcal{P}_i^0$ is simply replaced by the prior precision $\mathcal{I}_i^G$, yielding a straightforward prediction for how the decision maker incorporates prior information. On top of being biased, agents have a preference for familiarity: that is, they are more likely to select actions that they are better informed about ex-ante, as captured by the prior precision $\mathcal{I}_i^G$.

3.3 Discussion: Aversion to Prior Uncertainty and Revealed Preference

The choice probabilities in equation (3.11) load positively on the prior precision; in other words, agents display aversion to prior uncertainty. This begs the question of what feature of the model makes the agents uncertainty-averse.

The reason is to be found in the relationship between the functional form of their prior and the concavity of their utility function. According to the agent's prior beliefs, $\tilde{\Omega}_i$ follows a distribution with common mean and heterogeneous dispersion. If i was a risky lottery, $\tilde{\Omega}_i$ was the payoff of said lottery, and the prior G was an objective description of the distribution of lotteries' payoffs, a risk-neutral decision maker would be indifferent over which lottery i to pick.

However, what the agents maximize is $\tilde{\omega}_i$ (the logarithm of $\tilde{\Omega}_i$). Because the decision maker's utility is concave in $\tilde{\Omega}_i$, the reason why agents avoid choices with high prior dispersion is that they discount high states and strongly dislike low payoffs. It is costly to gather enough data to make low-precision choices (which are likely to yield highly-disliked low payoffs) become attractive.

One important implication of this solution is that, unlike under the standard multinomial logit, we can no longer infer payoffs from choice probabilities (revealed preference). To be more specific, suppose that we obtained data on choice probabilities $(\mathcal{P}_1(\boldsymbol{\omega}), \mathcal{P}_2(\boldsymbol{\omega}), ..., \mathcal{P}_n(\boldsymbol{\omega}))$. If (3.11) is the correct choice model, then (unlike in standard multinomial logit):

$$\mathcal{P}_i(\boldsymbol{\omega}) > \mathcal{P}_j(\boldsymbol{\omega}) \quad \not\Longrightarrow \quad \omega_i > \omega_j \tag{3.12}$$

In order to infer an ordering of payoffs from choice probabilities, we must have some degree of knowledge about the ordering of bias and prior precisions:

$$(\mathcal{P}_i(\boldsymbol{\omega}) > \mathcal{P}_j(\boldsymbol{\omega})) \ \wedge \ \left(\Delta_i \mathcal{I}_i^G \leq \Delta_j \mathcal{I}_j^G\right) \implies \omega_i > \omega_j \tag{3.13}$$

Similarly, we may be able to infer an ordering of prior precisions, as long as we can observe payoff orderings (as well as bias, if present):

$$(\mathcal{P}_i(\boldsymbol{\omega}) > \mathcal{P}_j(\boldsymbol{\omega})) \ \wedge \ (\omega_i + \delta_i \leq \omega_j + \delta_j) \implies \mathcal{I}_i^G > \mathcal{I}_j^G \tag{3.14}$$

This weakening of revealed preference has practical implications for empirical work, as well. Consider the following applied example.

Example. Suppose that the i actions are products, and that the decision maker is a consumer making a purchase. The consumer surplus from selecting product i depends on product prices (p_i), a vector of product characteristics ($\mathbf{x}_i$), as well as some "appeal" parameter α_i:

$$\omega_i \ = \ \alpha_i + \mathbf{x}_i' \beta - p_i \tag{3.15}$$

We can identify α_i as the intercept parameter by econometrically estimating equation (3.15) using data on prices, choices and product characteristics. If, however, the correct model is given by (3.11), the intercept parameter for product i will not identify α_i, but $\left(\alpha_i + \delta_i + \log \mathcal{I}_i^G\right)$. In order to restore the "appeal" interpretation of α_i, the econometrician may thus want to enrich the empirical model by adding additional explanatory variables (marketing, product age...) that can control for the consumers' prior knowledge of the products (i.e. that can proxy for $\delta_i + \log \mathcal{I}_i^G$).

To sum up, in settings where information is costly, the fact that some choice i is observed in the data with high frequency does not necessarily imply that decision makers rank it highly: it may equally reflect that the decision makers exogenously possess better information about that choice, or that their beliefs are biased. Thus, a key motivation for incorporating heterogeneity in the precision of prior information, is to be able to clearly separate (in theory as well as the empirics) preference orderings from prior information.

4 Application: Capital Allocation with Epistemic Uncertainty

In this section, I present an application of the RI-logit model with Tempered Stable prior. One of the less-known features of the RI-logit model, which will come useful in this application, is that it can be used to model risk as well as epistemic uncertainty.

In what follows, I show that the model can be deployed, with minimal modifications, to describe how risk-averse investors allocate capital to various economic activities in an economic environment that features risk as well as epistemic uncertainty. Epistemic uncertainty refers here to the lack of knowledge about the underlying parameters of the distribution of the assets' returns. This is distinct from "risk", which refers instead to the intrinsic randomness in the outcomes of the various available investments.

Let us consider the case where the n actions correspond to firms with risky technologies, which the agents can invest in. Consider a three-period model ($t = 1, 2, 3$). At time $t = 1$ a continuum of agents ι with mass one $\iota \in [0, 1]$ are born and endowed with a unit of capital. No consumption takes place in this period. In this period, agents acquire information about the payoffs from investing in different firms. In the second period ($t = 2$), agents choose one firm $i \in \{1, 2, ..., n\}$ in which to invest their unit of capital. In period $t = 3$, the agents collect a share of the consumption good produced by firm i and consumption takes place.

All firms produce the same homogeneous consumption good. The stochastic output of firm i is denoted by Y_i and is produced using the following technology, which is subject to lognormally-distributed productivity shocks with mean one:

$$Y_i \;=\; A_i K_i^{\beta} \cdot \exp\left(-\frac{\sigma_i^2}{2} + \sigma_i z_i\right) \qquad \text{with } z_i \sim N(0,1) \tag{4.1}$$

A_i is the (exogenous) deterministic component of productivity (with $a_i \overset{\text{def}}{=} \log A_i$), K_i is the (endogenous) capital stock, β is a returns to scale parameter, z_i is a (normalized) exogenous productivity shock and σ_i is the volatility of productivity. An agent who invests a unit of capital in firm i is entitled to a proportional share of the capital income. We call the return on that unit of capital invested R_i, and its expectation μ_i:

$$R_i \overset{\text{def}}{=} \frac{Y_i}{K_i} \qquad \text{and} \qquad \mu_i \overset{\text{def}}{=} \mathbb{E}(R_i) \;=\; A_i K_i^{\beta-1} \tag{4.2}$$

We assume that the shock z_i is pure risk: it cannot be learnt by the agent. On top of facing this risk, the agent faces epistemic uncertainty about the distribution of the payoffs. That is, the agents that populate this economy do not know the firms' technologies, and have to learn about the risk-return properties of each firm using the information acquisition problem described in Sections 2-3.

The agents have log utility over consumption, which makes them risk-averse. Also, their consumption is equal to the return on their investment, and there is no discounting (without loss of generality). Thus, at time $t = 2$, they choose a firm i in which to invest, by solving:

$$V_i(F) \;=\; \max_i \; \mathbb{E}^F(\log R_i) \tag{4.3}$$

Because the shock z_i is unlearnable, we can use the law of iterated expectations to restate this problem as:

$$V_i(F) \;=\; \max_i \; \mathbb{E}^F\left[\mathbb{E}\left(\log R_i | \mu_i, \sigma_i^2\right)\right] \;=\; \max_i \; \mathbb{E}^F(\tilde{\omega}_i) \tag{4.4}$$

$$\text{where} \qquad \tilde{\omega}_i \overset{\text{def}}{=} \log \mu_i - \frac{\sigma_i^2}{2} \tag{4.5}$$

The entries of $\tilde{\boldsymbol{\omega}}$ can now be interpreted as risk-adjusted expected returns. However, because of decreasing returns to scale, these are no longer exogenous objects: they are endogenously determined by the collective choice of all the agents in the economy (K_i) as well as by the exogenous productivities (A_i) and volatility parameters (σ_i).

Now, the previous literature has assumed, depending on the setting, that agents may acquire signals about exogenous states (in this case A_i, σ_i^2), collective actions (K_i), endogenous variables (μ_i) or a combination of the three (see Hébert and La'O, 2020, for a discussion). In general, what is appropriate to assume about the agents' ability to acquire signals depends on the specific setting.

In what follows, I assume that the agents can acquire signals about the risk-adjusted returns $\tilde{\boldsymbol{\omega}}$ (as opposed to A_i, σ_i^2 or K_i). I do so based on a "sufficient statistics" argument. Namely, ω_i is (from the point of view of an individual agent) a "sufficient statistic" for the expected payoff of action i. If the agent could know ω_i, there would be no additional value in knowing any of the other objects in the model (be them exogenous or endogenous).

Now, by restating the problem from (4.3) to (4.4), we have thus separated "risk" (uncertainty about the stochastic shock z_i) from "epistemic uncertainty" (lack of knowledge about the risk-adjusted expected returns ω_i). In this setting, agents can acquire information to reduce epistemic uncertainty, but not to reduce risk.

We now move back to $t = 1$. In this period, the agents solve the unrestricted information acquisition problem from equation (2.3). We assume that the agents' prior belief is that $\tilde{\Omega}_i$ follows a Tempered Stable distribution with mean $\mu_0 > 0$ and dispersion equal to $e^{\mathcal{U}_i}$.

Using Proposition 1, we obtain the following conditional choice probabilities:

$$\mathcal{P}_i(\boldsymbol{\mu}, \boldsymbol{\sigma}) = \frac{\mu_i \cdot \exp\left(-\frac{1}{2}\sigma_i^2 - \mathcal{U}_i\right)}{\sum_{j=1}^{n} \mu_j \cdot \exp\left(-\frac{1}{2}\sigma_j^2 - \mathcal{U}_j\right)} \tag{4.6}$$

A few interesting insights emerge from this equation. First, because each agent possesses one atom of capital and the total capital stock available to firms in period 3 is equal to one ($\sum_{i=1}^{n} K_i = 1$), this equation describes demand for risky assets in this simple economy. Specifically, asset demand is logit in expected returns and variances. This type of asset demand systems has been studied empirically by Koijen and Yogo (2019). Second, epistemic uncertainty about the distribution of returns makes asset demand *inelastic* (Gabaix and Koijen, 2021) – that is, the elasticity of demand for asset i with respect to its expected return is finite.

We next analyze how capital stocks and expected returns are determined in equilibrium, by providing a formal definition of the latter.

Definition (Equilibrium)**.** *An equilibrium is a vector of expected returns* $(\mu_1, \mu_2, ..., \mu_n)$ *and capital stocks* $(K_1, K_2, ..., K_n)$ *such that agents acquire information optimally at time 1 and invest optimally at time 2, the markets for capital clear – that is,* $\mathcal{P}_i(\boldsymbol{\mu}, \boldsymbol{\sigma}) = K_i$ *for all i – and expected returns are consistent with capital invested* (*equation 4.2*)*.*

Proposition 2. *The equilibrium expected rate of return to capital for firm i is given by:*

$$\log \mu_i = \frac{a_i}{2-\beta} + \frac{1-\beta}{2-\beta} \cdot \left(\frac{1}{2}\sigma_i^2 + \mathcal{U}_i\right) + \text{constant} \tag{4.7}$$

Proof. See Appendix. □

Further insights emerge from this result. First, as is typically the case in the presence of risk aversion, assets pay a risk premium; the risk premium is proportional the variance term $\left(\sigma_i^2\right)$. More interestingly, on top of the risk premium, assets also pay a premium for epistemic uncertainty ($\mathcal{U}_i$).

Another effect of information frictions is that expected returns co-vary positively with the fundamental productivity parameter a_i. This result might seem counterintuitive, at a first glance. To better understand it, let us consider what would happen if all assets had the same volatility $\left(\sigma_i^2 = \sigma^2\right)$ and agents had full information: market clearing would equalize the expected return to capital across companies, attaining a (constrained-)efficient allocation of capital across firms, and thus maximizing expected aggregate output ($Y = \sum_{i=1}^{n} \mathbb{E}(Y_i)$). Hence, under full information and symmetric risk, we do not expect to see any relationship between expected returns and physical productivity. In fact, we don't expect to see *any* dispersion in the rate of return to capital. Under imperfect information, agents are no longer able to discern the more profitable investment opportunity, and more productive firms (high A_i) receive too little capital. Because returns are decreasing in the capital stock, higher returns are a sign of inefficiently-low investment - this is well known from the previous literature on capital allocation (see David et al., 2016; Pellegrino et al., 2021). Hence, the fact that the expected returns co-vary with physical productivity reveals that information frictions lead to capital misallocation: agents provide too little capital to high-productivity firms, and too much to low-productivity firms.

To further investigate the implications of epistemic uncertainty for capital allocation, let us focus on the special case with symmetric risk $\left(\sigma_i^2 = \sigma^2 \ \forall i\right)$, and let us define aggregate output $Y = \sum_{i=1}^{n} \mathbb{E}(Y_i)$. Let us also define Y^*, the maximum level of output attainable by exogenously choosing the vector of capital stocks $(K_1, K_2, ..., K_n)$. We can then write the approximate loss in aggregate output, due to the information frictions, in terms of moments of the joint (cross-sectional) distribution of A_i and $\mathcal{U}_i$.

Proposition 3. *With symmetric risk, the percentage difference between the equilibrium aggregate output* (Y) *and the efficient aggregate output* (Y^*)*, due to information frictions, is equal* (*to a second-order Taylor Approximation*) to:

$$\frac{Y - Y^*}{Y^*} \approx -\beta \cdot \frac{\operatorname{var}(a_i) + (1-\beta)^2 \operatorname{var}(\mathcal{U}_i) + 2(1-\beta)\operatorname{cov}(a_i, \mathcal{U}_i)}{2(1-\beta)(2-\beta)^2} \tag{4.8}$$

where the variance and covariances weight firms by their respective shares in Y^**.*

Proof. See Appendix. □

This equation yields three additional insights with regard to how epistemic uncertainty affects the allocation of capital. First, the more dispersed is productivity (a_i) across firms, the more information frictions bite (if firms all have the same productivity, capital allocation decisions are inconsequential in the aggregate). Second, the more unequally-informed agents are (about firms), the more severe the misallocation of capital: "unfamiliar" firms command a premium on marginal returns, implying that they receive less capital than optimal, relative to more "familiar" firms. Third, the loss in aggregate output is positively related to the covariance between productivity and prior uncertainty: because capital fails to flow towards high-productivity and high-uncertainty firms, the extent to which these two groups overlap might either exacerbate or alleviate the cost of the information frictions (the latter occurs when high-productivity firms happen to be low-uncertainty firms).

5 Conclusion

In this paper, I provided a closed-form solution to the RI-logit model of Matějka and McKay (2015) that allows to study the impact of bias and prior uncertainty on choices probability. By assuming a Tempered Stable distribution for the agents' prior beliefs, I derived analytical formulae for the conditional and unconditional choices probabilities.

The solution obtained provides sharp predictions for how the agents' prior information affects their choice probabilities: ceteris paribus, the decision makers are more likely to select options that they are better informed about. In addition, they might have biased beliefs.

This formulation expands the applicability of the RI-logit to settings where prior information plays an important role: in particular, it allows researchers to study how (exogenous) changes in the prior information that is available to the agents affects decision making. Examples of markets where this modeling feature could be highly valuable include consumer finance, healthcare, and international capital markets.

I also provided an application of the analytically-solved model, in the form of a model of capital allocation where agents invest in risky firms subject to epistemic uncertainty. I derived implications for how epistemic uncertainty (which enters the agents' decision through the prior), impacts the elasticity of demand for assets, the expected return on various investments as well as allocative efficiency.

This paper contributes to the fast-growing methodological literature of rational inattention, initiated by Sims (2003), which is swiftly finding applications across several fields of economics, including macroeconomics, international finance, and industrial organization.

References

Brix, Anders. 1999. "Generalized gamma measures and shot-noise Cox processes." *Advances in Applied Probability* 31 (4): 929–953.

Brown, Zach Y, and Jihye Jeon. 2020. "Endogenous information and simplifying insurance choice."Technical report, Mimeo, University of Michigan.

Dasgupta, Kunal, and Jordi Mondria. 2018. "Inattentive importers." *Journal of International Economics* 112 150–165.

David, Joel M, Hugo A Hopenhayn, and Venky Venkateswaran. 2016. "Information, misallocation, and aggregate productivity." *The Quarterly Journal of Economics* 131 (2): 943–1005.

Gabaix, Xavier, and Ralph SJ Koijen. 2021. "In search of the origins of financial fluctuations: The inelastic markets hypothesis."Technical report, National Bureau of Economic Research.

Hébert, Benjamin M, and Jennifer La'O. 2020. "Information Acquisition, Efficiency, and Non-Fundamental Volatility."Technical report, National Bureau of Economic Research.

Jørgensen, Bent. 1987. "Exponential dispersion models." *Journal of the Royal Statistical Society: Series B (Methodological)* 49 (2): 127–145.

Koijen, Ralph SJ, and Motohiro Yogo. 2019. "A demand system approach to asset pricing." *Journal of Political Economy* 127 (4): 1475–1515.

Kolossiatis, Michalis, Jim E Griffin, and Mark FJ Steel. 2011. "Modeling overdispersion with the normalized tempered stable distribution." *Computational statistics & data analysis* 55 (7): 2288–2301.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt. 2021. "Rational inattention: A review."

Matějka, Filip, and Alisdair McKay. 2015. "Rational inattention to discrete choices: A new foundation for the multinomial logit model." *American Economic Review* 105 (1): 272–98.

McFadden, Daniel. 1973. "Conditional logit analysis of qualitative choice behavior." *Frontiers in Econometrics* 105–142.

Pellegrino, Bruno, Enrico Spolaore, and Romain Wacziarg. 2021. "Barriers to Global Capital Allocation."Technical report, National Bureau of Economic Research.

Shannon, Claude E. 1948. "A mathematical theory of communication." *The Bell system technical journal* 27 (3): 379–423.

Sims, Christopher A. 2003. "Implications of rational inattention." *Journal of monetary Economics* 50 (3): 665–690.

Steiner, Jakub, Colin Stewart, and Filip Matějka. 2017. "Rational inattention dynamics: Inertia and delay in decision-making." *Econometrica* 85 (2): 521–553.

Tweedie, Maurice C.K. 1984. "An index which distinguishes between some important exponential families." In *Statistics: Applications and New Directions - Proceedings of the Indian Statistical Institute Golden Jubilee International Conference*, Volume 579. 579–604.

Appendix (Proofs)

Proof to Proposition 1. By Lemma 2 in MM, (3.4) solves the RI-logit problem if and only if condition (3.6) holds. We therefore need to show that the latter is satisfied by:

$$\mathcal{P}_i^0 \;=\; \frac{\mathcal{I}_i^G}{\sum_{j=1}^n \mathcal{I}_j^G} \tag{A.1}$$

We guess and later verify that $\mathcal{P}_i^0 > 0$ for all $i = \{1, 2, ..., n\}$ - that is, all choices are selected with positive probability. This implies that (as pointed out in MM Appendix 1) condition (3.6) can be re-stated as:

$$\mathbb{E}^G \left(\frac{\tilde{\Omega}_i \, \mathcal{P}_i^0}{\sum_{j=1}^n \tilde{\Omega}_j \, \mathcal{P}_j^0} \right) \;=\; \mathcal{P}_i^0 \tag{A.2}$$

Next, note that, for any positive constant $\mathcal{C}$, we can use the scaling property of the Tweedie family of distributions to infer:

$$\tilde{\Omega}_i \;\overset{G}{\sim}\; \mathrm{Tw}_p \left(\mu, \frac{1}{\mathcal{I}_i^G} \right) \quad \Longrightarrow \quad \mathcal{C}\, \tilde{\Omega}_i \mathcal{I}_i^G \;\overset{G}{\sim}\; \mathrm{Tw}_p \left(\mathcal{C} \mu \, \mathcal{I}_i^G \,, \, \mathcal{C}^{2-p} \left(\mathcal{I}_i^G \right)^{1-p} \right) \tag{A.3}$$

Where Tw_p denotes the Tweedie-family distribution with power index p. Next, for $p \geq 2$, we can define the following parameters κ and θ:

$$\kappa \;\overset{\text{def}}{=}\; \frac{p-2}{p-1} \in [0,1] \qquad \text{and} \qquad \theta \;\overset{\text{def}}{=}\; 2\,(1-\kappa)\, \mu^{2-p} \tag{A.4}$$

then we can express μ in terms of θ as:

$$\mu \;\overset{\text{def}}{=}\; \left[\frac{\theta}{2\,(1-\kappa)} \right]^{\frac{1}{2-p}} \tag{A.5}$$

and for the following suitable choice of $\mathcal{C}$

$$\mathcal{C} \;=\; \left[2\,(1-\kappa)\, \theta^{1-p} \right]^{\frac{1}{2-p}} \tag{A.6}$$

we have that $\mathcal{C}\mu = \theta$ and therefore:

$$\mathcal{C}\, \tilde{\Omega}_i \, \mathcal{I}_i^G \;\overset{G}{\sim}\; \mathrm{Tw}_p \left(\theta\, \mathcal{I}_i^G, 2\,(1-\kappa) \left(\theta\, \mathcal{I}_i^G \right)^{1-p} \right) \tag{A.7}$$

For $p \geq 2$, this is what Kolossiatis, Griffin, and Steel (2011, henceforth KGS) call a Tempered Stable (TS) distribution. Jørgensen (1987, section 3) refers to these densities as Tweedie distributions “generated by positive stable distributions”. Brix (1999, section 2) refers to this subset of the Tweedie family as “Generalized Gamma” distributions. Despite the different names, this is always the same class of distributions.

Now, KGS follow a different parametrization (κ, δ, γ) from that of Tweedie-Jørgenson $\left(\mu, \sigma^2, p \right)$. The statement in (A.7) is equivalent to the following one, expressed using the notation of KGS:

$$\mathcal{C}\, \tilde{\Omega}_i \, \mathcal{I}_i^G \;\overset{G}{\sim}\; \mathrm{TS} \left(\kappa, \frac{\theta\, \mathcal{I}_i^G}{2\kappa}, 1 \right) \tag{A.8}$$

It can be easily verified that, regardless of whether we call this distribution Tweedie or Tempered Stable, the mean and variance are $\theta\mathcal{I}_i^G$ and $2\,(1-\kappa)\,\theta\,\mathcal{I}_i^G$, respectively. KGS provide expressions for the mean and the variance of the TS distribution in subsection 2.1 of their paper. This distribution becomes a Gamma distribution for $\kappa \to 0$ (or, equivalently, $p \to 2$). For $\kappa = \frac{1}{2}$ ($p = 3$) we have instead an Inverse Gaussian distribution.

The result in equation (A.8) implies, by Definition 4 of KGS, that:

$$\left[\frac{\tilde{\Omega}_1\,\mathcal{I}_1^G}{\sum_{j=1}^n \tilde{\Omega}_j\,\mathcal{I}_j^G}\,,\,\frac{\tilde{\Omega}_2\,\mathcal{I}_2^G}{\sum_{j=1}^n \tilde{\Omega}_j\,\mathcal{I}_j^G}\,,\,\dots\,,\,\frac{\tilde{\Omega}_n\,\mathcal{I}_n^G}{\sum_{j=1}^n \tilde{\Omega}_j\,\mathcal{I}_j^G}\right] \overset{G}{\sim} \text{MNTS}\left(\frac{\theta\,\mathcal{I}_1^G}{2},\frac{\theta\,\mathcal{I}_2^G}{2},\dots,\frac{\theta\,\mathcal{I}_n^G}{2};\kappa\right) \tag{A.9}$$

where MNTS denotes the *Multivariate Normalized Tempered Stable* probability distribution, a class of multivariate probability densities over the simplex that nests the Dirichlet distribution ($\kappa \to 0\ ,\ p \to 2$) as well as the Normalized Inverse Gaussian $\left(\kappa = \frac{1}{2}\ ,\ p = 3\right)$. The expected value of the marginal is derived in closed form by KGS in Theorem 1, Corollary 1. We can now verify that (A.1) respects (A.2) by plugging it in, and solving the expectation on the left hand side as the mean of a MNTS-distributed variable:

$$\mathbb{E}^G\left(\frac{\tilde{\Omega}_i\cdot\frac{\mathcal{I}_i^G}{\sum_{\iota=1}^N \mathcal{I}_\iota^G}}{\sum_{j=1}^n \tilde{\Omega}_j\cdot\frac{\mathcal{I}_j^G}{\sum_{\iota=1}^n \mathcal{I}_\iota^G}}\right) \;=\; \mathbb{E}^G\left(\frac{\tilde{\Omega}_i\,\mathcal{I}_i^G}{\sum_{j=1}^n \tilde{\Omega}_j\,\mathcal{I}_j^G}\right) \;=\; \frac{\frac{\theta}{2}\,\mathcal{I}_i^G}{\sum_{j=1}^n \frac{\theta}{2}\,\mathcal{I}_j^G} \;=\; \frac{\mathcal{I}_i^G}{\sum_{j=1}^n \mathcal{I}_j^G} \tag{A.10}$$

Because $\mathcal{I}_1^G > 0$ for all i (by assumption) it is thus verified that $\mathcal{P}_i^0 > 0$ for all $i = \{1, 2, ..., n\}$. □

Proof to Proposition 2. First, let us rewrite firm i's capital stock in terms of the expected return μ_i:

$$K_i = \left(\frac{\mu_i}{A_i}\right)^{-\frac{1}{1-\beta}} \tag{A.11}$$

This implies that the market clearing condition can be re-written as:

$$\left(\frac{\mu_i}{A_i}\right)^{-\frac{1}{1-\beta}} \propto \mu_i\cdot\exp\left(-\frac{1}{2}\sigma_i^2 - \mathcal{U}_i\right) \tag{A.12}$$

Rearranging:

$$\mu_i^{-\frac{1}{1-\beta}-1} \;\propto\; A_i^{-\frac{1}{1-\beta}}\cdot\exp\left(-\frac{\sigma_i^2}{2} - \mathcal{U}_i\right) \tag{A.13}$$

Taking logs and dividing by $\left(-\frac{2-\beta}{1-\beta}\right)$, we obtain equation (4.7). □

Proof to Proposition 3. Consider a second-order Taylor approximation of the change in aggregate output from its efficient level:

$$\Delta Y \quad\approx\quad \beta\mu^*\left[\sum_{i=1}^n \Delta K_i - \frac{1}{2}\sum_{i=1}^n \frac{1-\beta}{K_i^*}\left(\Delta K_i\right)^2\right] \tag{A.14}$$

This expression is derived using the fact that, absent risk, $\beta\mu^*$ is equal to the expected marginal product of capital and is equal across firms (at maximum output, expected returns are equalized). The fact that the total capital is fixed implies that the first-order term in parentheses is zero. We can then divide both sides by Y^* and rearrange the second-order term as:

$$\frac{\Delta Y}{Y^*} \approx -\frac{\beta}{2} \cdot \mu^* \sum_{i=1}^{n} (1-\beta) \frac{K_i^*}{Y^*} (\Delta \log K_i)^2 \tag{A.15}$$

We then use the following facts

$$\Delta \log \mu_i = -(1-\beta) \Delta \log K_i \tag{A.16}$$

$$\mu_i K_i = Y_i \tag{A.17}$$

to derive:

$$0 = \sum_{i=1}^{n} \Delta K_i = \sum_{i=1}^{n} \frac{\mu^* K_i^*}{1-\beta} \cdot \Delta \log \mu_i = \sum_{i=1}^{n} \frac{Y_i^*}{Y^*} \cdot \Delta \log \mu_i = \left(\sum_{i=1}^{n} \frac{Y_i^*}{Y^*} \cdot \Delta \log \mu_i \right)^2 \tag{A.18}$$

We finally plug equations (A.16) and (A.17) inside equation (A.15) to obtain:

$$\frac{\Delta Y}{Y^*} \approx -\frac{1}{2} \cdot \frac{\beta}{1-\beta} \sum_{i=1}^{n} \frac{Y_i^*}{Y^*} (\Delta \log \mu_i)^2 \tag{A.19}$$

Equation (A.18) implies that summation term in A.19 is the Y_i^*-weighted variance of $\log \mu_i$. By rewriting this variance in terms of the variance/covariance of $\log A_i$ and U_i we obtain equation (4.8). $\square$