Rational Inattention and Inference from Market Share Data*

Andrew Caplin, New York University, N.B.E.R.

John Leahy, New York University, N.B.E.R.

Filip Matějka†, CERGE-EI and CEPR

April 2016

Abstract

Following McFadden (1974), discrete choice models are typically estimated under the assumption of perfect information. We model discrete choice in a market with imperfect information. Each period new agents observe past market shares and, in the spirit of rational inattention, can acquire additional information of any selected form. Market shares are of generalized logit form, and reflect a mix of private and social learning. Simple formulae allow these to be separately identified. The logit form allows for application of standard empirical techniques, and correction of biases associated with the assumption of perfect information. We apply the identification strategy also to problems with self-selection, as in Heckman and Honoré (1990).

Keywords: discrete choice, information, rational inattention, market share, inference, self selection.

*We thank Dan Ackerberg, Michal Bauer, Sourav Bhattacharya, James Heckman, Štěpán Jurajda, Nobuhiro Kiyotaki, Nikolas Mittag, Juan Pablo Nicolini, Alessandro Pavan, Avner Shaked, Chris Tonetti, Xavier Vives, and Krešo Žigič for helpful discussions.

†This research was funded by GAČR grant P402-14-30724S

1 Introduction

Economists analyze discrete choice in such diverse settings as school selection, job selection, choice of geographic location, choice of health insurance plan, and choice of political party affiliation. Following the pioneering work Block and Marschak (1960) who introduced the random utility model, it was the logit model of McFadden (1974) that opened these floodgates. This workhorse model continues to be further enhanced and estimated in ever richer settings to this day.

Well-developed as is the current approach, it involves a restrictive assumption: that information is perfect. Choice probabilities depend only on preferences over the goods as the econometrician defines them, not on subjective beliefs. While this may be valid for simple repeated choices such as mode of transport, it is far less credible for choice among such complex products as insurance contracts, let alone the choice of a job. Interestingly, this concern was first noted by Block and Marschak themselves, who felt that simple data on stochastic choice would be inadequate to separate fluctuations in "perceptibility and desirability". This issue is even more general. Preferences and beliefs are the most universal elements in economic modeling. It is the fact that these are not directly observed that forces us to make extreme assumptions.

We show in this paper that rational inattention theory of Sims (1998, 2003) provides useful methods for incorporating both preference heterogeneity and imperfect information into models of discrete choice. We develop the simplest framework for market analysis that renders this work operational. Consumers enter a market with a discrete set of available alternatives. These alternatives are heterogenous and different consumers value them differently. The problem that consumers face is that they have imperfect information concerning which option is best for them. Before they enter the market, they have some prior beliefs regarding the distribution of matches. They also learn from two sources. As in Becker (1981) and Caminal and Vives (1996), they freely observe past market shares. This allows them to learn from the behavior of past entrants. They can also choose to acquire additional private information about the options. It is this individual learning that gives past market shares their informational content. This additional information is costly to process. Following Sims (2010) and Woodford (2009), the cost of information is linear in the Shannon mutual information between prior and posterior beliefs.

We allow for general heterogeneity in agents' preferences, any finite number of choice alternatives, and a general form of attentional effort. In spite of this generality, the resulting model is very tractable. The source of this tractability lies in the fact that the observable market shares end up strongly constraining the effect that consumers' prior beliefs have on their choice. Left unrestricted, prior beliefs could justify any action. A key result of our analysis is that steady state market shares of chosen products are precisely those that would follow from full knowledge of the true distribution of preferences in the population. Thus, these shares can be computed using only the distribution of preferences and not initial beliefs.

There is a natural feedback between choice and beliefs in our model. This complementarity exacerbates the effect of market share on choice. Precisely because market shares reflect the underlying heterogeneity in preferences, they are useful in guiding individual attention and choice. A popular item is therefore more likely to be chosen by agents of all types. This is sensible. The attention garnered by books, films, and restaurants that achieve popularity raises further interest in them (as in Moretti 2010). The fact that some books are in New York Times Best Sellers and some are not affects and guides the form additional private attention across all books.

The resulting model is a reinterpretation of the classic logit model. Following Matějka and McKay (2015), we know that modeling information acquisition in the way that we do gives rise to choice probabilities that have a generalized logit form in which the payoffs to an action are weighted by the probability that an action is chosen. Our model links these probabilities to market shares and market shares to preferences. This allows us to separately identify the effects of perceptibility and desirability. This analogy with the logit model is very convenient and allows for empirical work with standard techniques. There are two important differences between our model and the standard discrete choice model. The first is the interpretation of the error term. In the standard model, choice appears stochastic because utility is stochastic. In our model, choice is stochastic because learning is stochastic. Agents in our model make mistakes. They occasionally choose options that they would not choose if they had better information. The second difference is the interpretation of the payoffs. In our model, higher market share, by increasing the perceived desirability of an option, enters the model in exactly the same way as an increase in utility, thereby biasing estimates of utility in models that ignore attention frictions. While of generalized logit form, our model also allows for realistic failures of the controversial independence of irrelevant alternatives (IIA) feature of the standard formulation. Thus it may in some cases be a valuable alternative to nested-logit models and other models designed precisely to by-pass this limitation of the McFadden model.

Despite the heterogeneity in true preferences, our model has simple welfare properties.

Since our agents choose as if they know the true distribution of preferences over chosen options, they do the best they can given the available alternatives. The only long run failure of social optimality arises when a potentially popular option remain unchosen. With zero market share past choice provides no positive signals that would overcome an inappropriately pessimistic prior. In this way the model focuses attention on policies that would affect the set of available options.

In comparison with the prior literature on rational inattention, this paper takes us a step closer to application. For example, while the logit form of demand was known (Matějka and McKay 2015), it depends on unobservable prior beliefs. In the current paper, the observable market shares appropriately summarize the effect of prior beliefs. It is this feature that allows us to disentangle the effects of preferences and beliefs on choice. Other theoretical models of rational inattention include Luo (2008), Woodford (2009), Mackowiak, Wiederholt (2009), Van Nieuwerburgh and Veldkamp (2010), Mondria (2010), Matějka (2010), and Caplin, Dean and Leahy (2016). To date there has been limited corresponding empirical work (Mondria et al. (2010); Bartoš et al. (2015)). The results in this paper are designed to speed the model's integration into discrete choice analysis.

To further highlight links with application, we show that our model allows for identification of the distribution of types in a classical self-selection problem of Heckman and Honoré (1990). The identification is possible if the econometrician has access to the same data as in their perfect information case. This finding emphasizes the applicability of the transparent interaction of beliefs and preferences in our logit formula, and of the role of market shares. We show also that our model allows in principle for precise identification of individual preferences. In our model rich data allows one to identify preferences of all types for all alternatives in the market.

Our work relates to the broader literature on social learning. In our model, the key role of high market share is that it guides buyers' attention towards such high selling items, while other items might be ignored by incoming consumers altogether. This emphasis on social learning is important in many discrete choice settings. It has been identified in such diverse settings as: technology adoption (e.g. Munshi (2003)); retirement saving plan choice (Duflo and Saez (2002)); health insurance choice (Sorensen (2005)), regular product choice (Moretti (2010)); and choice in restaurants (Cai, et al. (2009)). There is also a related literature in political science where it is common practice to infer political preferences from opinion polls and vote share. In this setting the influence of differences in information are increasingly under investigation (Bartels 1996, Delli, Carpini and Keeter 1996). Our model provides new ways to consider these frictions in inference as well. Finally, our consumers make a once and for all choice. There is another literature that considers learning from experience. Erdem and Keane (1996), Ackerberg (2003) and Osborne (2011) consider how to identify beliefs and utility in such settings. Future work would combine learning from market share with learning from experience.

The model is introduced in the next section, and Section 3 presents the main theoretical results. Implications for applied work are presented in Section 4. Section 5 illustrates basic features of the model and its connection to empirical work by discussing a simple example, and Section 6 concludes.

2 Model

There is a fixed finite set $A = \{1, \dots, N\}$ of options. Time is discrete and indexed by $t \geq 0$. Each period a continuum of new agents make a once off choice from A. Agents are of a finite number of distinct preference types $\omega \in \Omega$, and there is an underlying utility function,

$$u : A \times \Omega \to \mathbb{R}.$$

The agents do not know the utilities from selecting different options. Agents, however, can refine their beliefs about ω in two ways. First, they observe choice frequencies of all prior generations, i.e. all past market shares. Second, the agents can in the style of rational inattention process additional costly information about ω, and thus about the utilities, by exploring the offered options in more detail.

Let $\Delta(\Omega)$ denote the set of densities over types and let $g^* \in \Delta(\Omega)$ denote the true density of preference types in the population. $g^*(\omega)$ is then the share of type ω agents. g^* is fixed and does not change over time. We place no restriction on the form of the utility function and how it varies across goods and types.

The agents are Bayesian. In period t, each agent first observes past market shares. They then enters the choice situation with a common prior belief $\mu_t \in \Delta(\Omega)$, where $\mu_t(\omega)$ is the probability that the agent is of type ω.

Agents are also rationally inattentive. In addition to relying on the information from past market shares, each agent can obtain additional information about ω, and thus about the utilities of available options, by exploring the offered options in more detail. In practice, this might involve personal examination of an alternative such as test driving a car or a visit to a store; it might involve a detailed reading of the product reviews in Amazon.com or in yelp.com; or it might involve discussions with friends, colleagues, or other people that the agent regards as similar to him or herself. We do not place structure on the process of this private information acquisition. Rather we allow agents to choose an arbitrary information structure subject to a cost of choosing more informative structures. This means that agents can process information in any way they like. They may choose to learn more about the utility they derive from several of the available options, or work to compare particular pairs of products, etc.

To model the cost of information acquisition we employ the standard model of rational inattention.[1] The agents' choice of what pieces of information to acquire and how to respond to such information can be described in a compact way by the distribution of choices conditional on the unknown type ω and the prior μ,[2]

$$P(i|\omega) = \Pr\{i \in A|\omega \in \Omega, \mu\}. \tag{1}$$

Given a prior belief μ, the agent maximizes:

$$V(A,\mu) \;=\; \max_{\{P(i|\omega)\}_{i\in A,\omega\in\Omega}} \sum_{\omega\in\Omega} \mu(\omega) \left(\sum_{i\in A} P(i,\omega|\mu) u(i,\omega) \right) - \lambda I(P). \tag{2}$$

The first term on the right-hand side is the expected utility of the strategy $\{P(i|\omega)\}_{i\in A,\omega\in\Omega}$. The second term is the cost of information acquisition which we take as proportional to the mutual information between the choice i and the type ω, as in Sims (2003). $\lambda > 0$ is the marginal cost of information. Mutual information I measures the expected reduction in entropy of beliefs over ω.

$$I(P) = -\sum_{i\in A} P(i) \ln P(i) + \sum_{\omega\in\Omega} \mu(\omega) \left(\sum_{i\in A} P(i|\omega) \ln P(i|\omega) \right), \tag{3}$$

[1]See Matějka and McKay (2015) and Caplin, Dean, and Leahy (2016)

[2]To understand this formulation, note that an information strategy is described by a joint distribution of types and signals, which defines how likely are agents of each type to receive each signal. In terms of its impact on action choice, one can characterize each signal by the corresponding posterior belief that it induces. Upon acquiring each particular signal and identifying the corresponding posterior belief, the agent chooses among available options to maximize expected utility. Therefore, signals can be associated with corresponding optimal choices. Since more Blackwell informative information structures are more costly, no option will be chosen from two distinct signals. Hence a strategy of information acquisition is equivalent to a type-dependent stochastic choice function $P(i|\omega)$. Each such function specifies the probability of observing information that prompts the choice of action i when the agent's true type is ω and the prior is μ.

where $P(i) \equiv \sum_{\omega \in \Omega} \mu(\omega) P(i|\omega)$ is the average probability of choosing i conditional on the prior μ only. The RHS in (3) can be reorganized to equal the entropy of prior beliefs less the expected entropy of posteriors. If the agent does not acquire any information about ω, then the cost is zero, and more information is more costly. Given the concavity of $\ln P(i|\omega)$, it is increasingly costly to make $P(i|\omega)$ type-contingent. This specific entropy-based cost of information allows for tractability. Moreover, it can be derived from not unreasonable axioms (Csiszár (2008)) and can be microfounded using fairly general assumptions on the technology of information acquisition (Cover and Thomas, 2006).

Finally, the type dependent choice probabilities $P(i|\omega)$, along with the true density of types g^*, generate the realized market shares $M(i)$. We assume that private information acquisition is conditionally independent across agents, which implies:

$$M(i) = \sum_{\omega \in \Omega} g^*(\omega) P(i|\omega). \tag{4}$$

3 Choice behavior and market shares

The first order condition of the objective (2) implies[3]

$$P(i|\omega) = \frac{P(i) e^{u(i,\omega)/\lambda}}{\sum_{j \in A} P(j) e^{u(j,\omega)/\lambda}}. \tag{5}$$

The agents' choice is stochastic, and follows a logit model. The more utility a particular option provides, the more likely it is that the agent selects it, and this dependence is stronger for lower costs of information λ. The noise in the choice is driven by noise in signals that the agent receives.

The only difference between (5) and the standard form of the logit model is that in the standard model only the utilities appear in the exponents, but here the expression also includes $P(i)$. This is the ex ante probability of choosing i, conditional on the prior only, and not on the true type. This probability describes what the agents expect before they start collecting any private information. If the cost of information is positive, then prior knowledge affects the agent's choice via two channels. First, simply because a Bayesian agent weights both noisy signals as well as prior beliefs, and second, the prior knowledge also affects the way the agent chooses to process information. The more likely it is that an

[3]See Matějka and McKay (2015).

agent chooses a given option ex ante, the more bias there is towards the option even if the true utility from it is low.

The particular logistic functional form is driven by the entropy-based cost, which determines the optimal form of signals. These signals maximize expected utility for a given cost of information, i.e, for given entropy of posterior beliefs.

While it has been known that rational inattention delivers the decision model (5), one drawback is that there is no general closed-form formula for the ex ante probabilities $P(i)$. More importantly, from an applied perspective, while the behavior (5) describes the interaction of preferences, beliefs and attention choices in an elegant way, the unconditional probabilities $P(i)$ depend on the prior beliefs, which are unobservable. The question then is how to disentangle preferences from beliefs given some observed choice behavior.

The model enables us to identify $P(i)$ from market share data. We do this by showing that the market converges to a steady state, where market shares reveal the unconditional probabilities $P(i)$.

Lemma 1. *Agents' beliefs μ_t converge to a steady state after a finite number of periods.*

Appendix A details how the priors of different generations of agents evolves over time. All proofs are contained in Appendix B.

The idea behind the proof is that agents in period t have a set Γ_t of possible distributions of types g. What drives convergence is a disconnect between the expected probability of choosing an option and the observed market shares. A distribution g from this set Γ_t is eliminated if the resulting market shares could not be generated by g, i.e., if

$$M(i) \neq \sum_{\omega \in \Omega} g(\omega) P(i|\omega).$$

(4) provides orthogonality conditions, and the finite dimension of Ω guarantees that the model converges to a steady state prior belief in a finite number of periods, when all distributions g in the limit set give rise to the observed market shares. Note that while the limit set includes the true g^*, it might also include other distributions of types. Agents are not guaranteed to learn g^* perfectly.

The next proposition states that the choice probabilities $P(i)$ can be identified in steady state: they equal the market shares. Intuitively, if this were not the case, then convergence of beliefs would not yet be complete.

Proposition 1. *Steady state market shares satisfy:*

$$M(i) = P(i), \tag{6}$$

and the choice behavior (5) *takes the form:*

$$P(i|\omega) = \frac{e^{u(i,\omega)/\lambda + logM(i)}}{\sum_j e^{u(j,\omega)/\lambda + logM(j)}}. \tag{7}$$

The next proposition links these market shares to choice given the true distribution g^*. Now, the choice behavior under rational inattention is fully characterized by agents' preferences, unlike in the previous work. The effect of prior beliefs is pinned down by the observed market shares which themselves are a function of choice.

Proposition 2. *In the steady state, $P(i)$ solves the rational inattention problem* (2) *for the prior $\mu = g^*$ and a given set of selected options $\bar{A}$.*

We call this the *"as if"* result. Even though agents may not know the true distribution of types, their resulting choices are just as they would be if they did know it. The reason is again that otherwise there would be a disconnect between the observed market shares and the known strategy $P(i)$, and convergence would not be complete. The statement holds only when the choice set is not extended beyond the set of selected options in steady state. This is because some options may not be selected in steady state. Given prior beliefs, there may be options that are never chosen, so that agents never learn about them. Yet full information may reveal that these should be very popular options. It is only when we conditioning on the selected options that $\mu = g^*$ delivers the exact steady state $P(i)$ and also market shares.

This result helps us out in two ways. First, the fact that agents act as if they know the true distribution of types in steady state greatly simplifies the analysis of the model and limits the range of steady state behavior. If one knows the steady state choice set $\bar{A}$, one can always assume that agents know the true distribution of types. We will use this in the following sections. Second, the result implies that market inefficiency takes a very limited form, which we discuss in the following section. We show below that the only welfare inefficiency that can arise in the steady state relates to good options that are not selected at all. Conditional on the set of options with positive market shares, the market shares are social optimal.

4 Inference from Market Share Data

This section addresses the essential issue of how the model relates to data on market shares. We show that the results in Section 3 are useful for identification purposes, and also for specification tests. We first address the implications of our model for applied work on discrete choice in the tradition of McFadden (1974). In that literature noise in choice is driven exclusively by heterogeneity in taste, i.e. random utility, while in our model it is also driven by noise in signals. We discuss how the findings impact inference in cases when the attention frictions are present. We then show that our model allows for identification of the distribution of types in a variant of the self-selection problem of Heckman and Honoré (1990) with imperfect information. The identification is possible if the econometrician has access to the same data as in the perfect information case of Heckman and Honoré (1990). This finding emphasizes the applicability of the transparent interaction of beliefs and preferences in (5), and of the role of market shares in Proposition 1. The simple formulae the model produces stand out in this application: in a general model of private and social learning, confounding of agents' heterogeneous preferences and beliefs would make identification far more challenging. We close the section by showing that our model allows an econometrician with ideal data to precisely identify utility functions of all types for all purchased products. This is because these preferences pin down the distribution of type-specific choices. This inference is far richer than would be possible if agents had perfect information, since in this case one would only observe the best option for each type.

4.1 Connection to random utility models

Our findings relate to a long tradition in applied work of using market shares to infer utility parameters. Prominent examples include McFadden (1974) and Berry, Levinsohn and Pakes (1995), but this literature is much larger. Let us first discuss the close connection to McFadden (1974). In this seminal paper, an agent n of type ω derives utility

$$U_{i,n}^{\omega} = U_i^{\omega} + \beta\epsilon_{i,n} \tag{8}$$

from choosing an option i, where U_i^{ω} is the intrinsic utility from the option i, while $\epsilon_{i,n}$ accounts for the unobserved heterogeneity among agents, and $\beta > 0$ is a scalar. Here $\epsilon_{i,n}$ is a taste shock. If $\epsilon_{i,n}$ is extreme value distributed, then conditional on the observables, the type dependent market share is given by the logit model,

$$M(i|\omega) = \frac{e^{\beta U_i^\omega}}{\sum_j e^{\beta U_j^\omega}}. \tag{9}$$

The subsequent literature then proceeds to estimate U_i^ω, or imposes additional structure on the utility, and the goal is to estimate how it depends on characteristics of the options.

In our model, due to Proposition 1, the type dependent choice in steady state follows instead:

$$M(i|\omega) = \frac{e^{u(i,\omega)/\lambda + logM(i)}}{\sum_j e^{u(j,\omega)/\lambda + logM(j)}}. \tag{10}$$

Here we assume conditional independence of signals, which implies $M(i|\omega) = P(i|\omega)$.

Our model provides an alternative interpretation of stochastic choice data, as opposed to the randomness driven by unobserved heterogeneity. In this model, the randomness in choices is instead driven by the randomness in choice mistakes due to the noise in signals.

Our model suggests a potential bias in inference of utility parameters if the information frictions are present, but neglected. The formulas (8)-(10) imply the following proposition.

Proposition 3. *Cross-sectional choices in our model are observationally equivalent to choices in random-utility model with*

$$U_{i,n}^\omega = \Big(U_i^\omega + \lambda logM(i)\Big) + \frac{1}{\lambda}\tilde{\epsilon}_{i,n}, \tag{11}$$

where $\tilde{\epsilon}_{i,n}$ is extreme-value distributed.

Therefore, if we use the standard logit model for inference of utility from choice data, what we estimate is $(U_i^\omega + \lambda logM(i))$ instead of U_i^ω. In our model, high market share is self-enforcing, and thus the effects of characteristics that are associated with options with high market shares will be biased upward and those with low market shares are biased downward. Of course, to infer U_i^ω, econometricians must first estimate $\frac{1}{\lambda}$ in precisely the way that β is currently estimated, and then subtract $\lambda logM(i)$, which is observed.

More generally, the findings above imply that if there is both unobserved heterogeneity $\epsilon_{i,n}$ as well as noise in attention $\tilde{\epsilon}_{i,n}$ present, then the observationally equivalent random utility model is

$$U_{i,n}^\omega = \Big(U_i^\omega + \lambda logM(i)\Big) + \frac{1}{\lambda}\tilde{\epsilon}_{i,n} + \beta\epsilon_{i,n}. \tag{12}$$

For a general distribution $\epsilon_{i,n}$ the choice is not necessarily logistic. However, the implication of the subtraction of $\lambda logM(i)$ from utilities inferred without considerations of the attention friction holds. See Cardell (1997) for description of distributions that preserve the logit property.[4]

Note that the testable implications of the type-dependent choice (10) differ from the standard logit (9). Most importantly, (9) satisfies the independence of irrelevant alternatives (IIA) axiom. Let i, j be two existing alternatives with positive market shares. If a new alternative is added to the choice set, then the ratio of type-dependent market shares of the two original alternatives remains unchanged,

$$\frac{M(i|\omega)}{M(j|\omega)} = \frac{M^*(i|\omega)}{M^*(j|\omega)},$$

where the asterisk denotes market shares for the expanded choice set. This property of the logit model is often seen as unrealistic, as in the red bus-blue bus problem of Debreu (1960). If an agent chooses between a bus and a train, both with the probability 1/2, and a duplicate bus is added to the choice set, then according to the logit model, the resulting probabilities should be $(1/3, 1/3, 1/3)$. Debreu argued that this results is unappealing since the addition of the second, identical bus should mainly affect the probability of choosing the first bus, and the choice probabilities should rather be 1/2 for the train, and 1/4 for each of the buses. The literature on nested-logit, and similar models, addresses these concerns by explicitly accounting for correlations of taste shocks across options.

In our model, instead, a change in the choice set will alter market shares. Proposition 1 immediately implies the following proposition.

Proposition 4. *In our model, type-dependent market shares satisfy*

$$\frac{M(i|\omega)}{M(j|\omega)} = \left(\frac{M(i)M^*(j)}{M(j)M^*(i)}\right)\frac{M^*(i|\omega)}{M^*(j|\omega)}. \tag{13}$$

IIA is broken if a change of the choice set implies

$$\frac{M(i)M^*(j)}{M(j)M^*(i)} \neq 1.$$

Since M and M^* are observable, then (13) allows for testing of the model using precisely the same methods as for the standard logit model. In case of the Debreu (1960) problem, our model would imply that the addition of the new bus decreases the unconditional market-

[4]See also Dasgupta and Mondria (2012) for an application of Cardell's findings to international trade with rational inattention.

share of the original bus to 1/4, while the market share of the train would remain unchanged. IIA is broken, but the choice still follows the market-share-augmented logit model (10).

In our model, IIA does not hold, but it breaks in a very specific way, which still allows the model to be tractable and estimation feasible. Instead of using an explicit and a more complex functional form, e.g., nested logit, our model implies that all correlations in preferences across options and agents translate into the market shares $M(i)$, and affect the type-dependent choice via (10). From a theoretical perspective, the exact market shares that emerge for a given choice set are given by Proposition 2.

Finally, let us comment on inference of elasticities. Consider the specification in Berry, Levinsohn and Pakes (1995),

$$U_i^\omega = X^i \gamma_\omega - \alpha p_i, \tag{14}$$

where X^i is the vector of characteristics of an alternative i and γ_ω is the vector of utility elasticities, α is the price elasticity and p_i is the price of good i. In our model, a rise in price p_i tends to reduce $P(i)$ and thus the market share $M(i)$, which further reduces the individual demand via the learning channel. Because of the learning, data on demand tends to overstate the price elasticity α. The same effects are present for the other elasticities in γ_ω.

4.2 Inference of skill distribution: self-selection and identification

In this section we follow the work of Heckman and Honoré (1990). The general issue of interest there is that econometricians observe agents who have potential outcomes in multiple states of the world that are subject to self selection (e.g., wages when participating in a program or not, or with a specific level of education). The intricacy is that the actual state that agents choose is endogenous. Moreover, the state they choose is not random. Typically it is modeled as being their best alternative from the choice set. An econometrician who aspires to assess a policy of shifting agents from one state to another needs to know the joint distribution of outcomes. The question is under which conditions the underlying joint distribution can be inferred from the realized and selected distributions.

In their seminal work, Heckman and Honoré (1990) study identification of the distribution of skills from the distributions of wages. Thus, while the econometrician observes wages in sectors that the workers chose to work in, data on unrealized wages in other sectors is not available. They show that observation of cross-sectional data on wages is not sufficient to identify the skill distributions if the class of skill distributions extends beyond log-normal distributions. However, they find that sufficiently rich panel data (cross-sections for various sectoral wages) allows for the identification.[5] In this section we adapt our model to this setup. In that manner we consider self selection in which agents who are imperfectly informed about their skills use a mix of market data and personal investigation to choose between sectors. In economic terms, this seems like a valid generalization of their model.

Heckman and Honoré (1990) consider two sectors, $i \in \{1, 2\}$, and a given distribution $g^*(\omega)$ of types of agents. The type $\omega = (s_1, s_2)$ is given by skills s_i in the two sectors. In sector i, there is a unit wage W_i, and if an agent chooses to work in sector i, an agent receives a wage $Y_i = W_i s_i$. Agents always choose the sector that provides them with a higher wage. The econometrician observes the distribution of W_i for all i only, and the task is to identify the distribution of $\omega = (s_1, s_2)$.

To adapt our model to accommodate this structure, we correspondingly define the utility function as depending on the sector and skill according to the corresponding wage,

$$u(i, \omega) = W_i s_i.$$

The key change in the model is that our agents do not always choose the sector that pays better, since they do not know their type. They may potentially make mistakes in their occupational choices because they do not know ex ante how well they would do in a sector, or how valued they would be by firms in the sector. They observe the unit wages w_i in each sector, but do not know their personal level of skill. We assume that there is a finite number of different types ω,[6] and each type is represented by a large number of agents. Let $g_i^*(s_i)$ denote the share of agents working in sector i with skill s_i,

$$g_i^*(s_i) = \sum_{s_j; j \neq i} P(i|\omega) g^*(\omega) \quad s.t. \quad \omega = (s_1, s_2).$$

Note that the model fits our general framework. Hence proposition 1 implies that in steady state,

$$P(i|\omega) = \frac{M(i) e^{W_i s_i / \lambda}}{\sum_{j \in \{1,2\}} M(j) e^{W_j s_j / \lambda}},$$

[5]By observing how distributions of wages across sectors respond to different wage profiles across sectors, econometrician can recover how many workers there are of each type, where a type is a particular combination of skills across sectors.

[6]This assumption could be relaxed at the expense of notational complexity, since the model would then need to be formulated using probability measures instead.

where $M(i)$ denotes the share of workers choosing sector i. As in Heckman and Honoré (1990), we assume that the econometrician has access to panel data on market shares and on the cross-sectional distribution of wages in each sector for a selected set of unit-wage pairs (W_1, W_2). The key result is the natural generalization of theirs: that this rich data allows the econometrician also to recover g_i^* for each of these pairs, since $s_i = Y_i/W_i$.

Proposition 5. *Panel data on market shares and sector specific wage distributions generically identifies the distribution of skills g^* and the cost λ.*

That our model allows for identification is a direct consequence of the fact that market shares summarize the effect of learning for individual's behavior in an elegant way. Conditional on observing the market share, the econometrician knows $P(i)$ even if the underlying distribution of types remains uncertain.

4.3 Inference of individual preferences

The rational inattention model introduces a non-standard information asymmetry. An outside observer with access to suitably rich data on market shares may be better able to understand preferences than are decision makers themselves. Agents in the model are acquiring information optimally given their limited resources. They focus their attention on matters that concern them directly, but their powers are limited. One could imagine that a large agent, such as the government, Google, Amazon, or a market research firm such as J. D. Power and Associates or Consumer Reports, might have greater access to large amounts of detailed choice data as well as greater incentives to process this information. Governments and research firms collect a broad range of statistics. Google sees the search behavior of a large fraction of agents. Amazon directly observes consumer choice. In this era of "big data" such an agent might be able to put together detailed market data and might be able to learn type-dependent market shares $M(i|\omega)$.

The following proposition shows how to use type-dependent market shares to recover preferences; this is a follow-up on Section 4.1. A simple ratio test allows one to infer agents' preferences and reveals optimal choices by type. Contrary to other models, in our model the whole set of preferences across all alternatives is identified, while many other models allow for identification of the top alternative only, since that is the only one agents choose.

Proposition 6. *In the steady state,*

$$\frac{u(i,\omega) - u(j,\omega)}{\lambda} = \log\left(\frac{M(i|\omega)}{M(j|\omega)} \Big/ \frac{M(i)}{M(j)} \right). \tag{15}$$

The proposition follows almost immediately from Proposition 1. Note that all expected utilities can be identified up to the cost of information λ, and individual choices under different aggregate market shares are identified fully. Notice also that this inference does not depend on knowledge of the agents' beliefs or of the distribution of types in the population g^*.

Proposition 6 implies that an outside observer can infer preferences from detailed market share data. Learning from market share skews choice in the direction of popular choices, and for this reason popular choices tend to be popular for all types. That being said, optimal private attention also skews choices in the direction of individual payoffs. One can infer whether an agent of type ω prefers good i to good j by comparing the frequency with which agents of type ω choose these goods to the average frequency of purchase in the population. Even if an agent of type ω chooses option i very rarely and option j quite frequently, if they choose i relatively more frequently than does the average agent, one can infer that they in fact prefer option i to option j. This holds as long as both options are chosen by some agents in the general population, i.e., $M(i) > 0, M(j) > 0$.

In many other models, e.g., with optimal deterministic choice, type-specific choice would not be very informative as it would reveal the most preferred option only. Thus if the choice set does not vary, then the econometrician can infer what option is preferred by a particular type, but does not learn anything else about utilities from less preferred alternatives. Here, on the other hand, the choice is probabilistic with probabilities reflecting underlying expected utilities. Type-dependent choices reveal not only the most preferred option, but the entire ranking of observed choices. The simple ratio test is particularly striking. More general models of social and private learning would produce far more intricate inference, and it would be hard if not impossible to carry out the equivalent identification exercise.

5 Example

We now solve for the full evolution of the simplest market with two types and two choices, and illustrate its implication for inference of utilities from data. Suppose $u(i, \omega)$ is such that type 1 prefers option 1 while type 2 prefers option 2,

$$u(1,1) > u(2,1) = 0 = u(2,2) > u(1,2).$$

Solving (2) for a given prior μ, we immediately get the following.[7] Defining,

$$\tilde{P}_1 = \frac{\mu(1)}{1 - \exp(u(1,2)/\lambda)} - \frac{1 - \mu(1)}{\exp(u(1,1)/\lambda) - 1}, \tag{17}$$

the unconditional probability of choosing the first option satisfies,

$$P(1) = \begin{cases} \tilde{P}_1 & \text{if } \tilde{P}_1 \in [0,1] \\ 0 & \text{if } \tilde{P}_1 < 0 \\ 1 & \text{if } \tilde{P}_1 > 0 \end{cases} . \tag{18}$$

According to (5), all type dependent choices $P(i|\omega)$ are determined by the unconditional choice probabilities $P(i)$, the payoffs $u(i,\omega)$ and the cost of information λ.

In the two-by-two case, the market converges to steady state in either period 0 or period 1. If prior beliefs are such that only one option is chosen in period 0, agents learn nothing about the population from market share, and the period 0 choices repeat themselves in subsequent periods. If both options are chosen in period 0, then the market reaches steady state in period 1. This is because period 0 market shares perfectly reveal the share of type 1 agents: (5) implies that type 1 agents are more likely to choose option 1 than type 2 and these choice probabilities are known. By Propositions 1 and 2, the steady state market shares are given by (18) for $M(i) = P(i)$ and with a prior that equals the true distribution of types, i.e., $\mu = g^*$.

The two-by-two case has well-behaved comparative statics. First, making one option more appealing for one type increases the probability that any type chooses it. This is because the type for whom the option becomes more appealing pushes the market share up, which via inference from market share implies that the other type is also more likely to choose it. This is easy to see from (17), where $\tilde{P}_1$ is increasing in both $u(1,2)$ and $u(1,1)$. Second, an increase in $g^*(1)$, and thus a steady-state $\mu(1)$, will also cause the probability of choice 1 to rise. Finally, as information costs λ rise, the influence of the ex ante optimal choice grows, and the interaction between agents becomes stronger. As λ rises, agents tend to rely more and more on their prior. Eventually, λ rises so high that $P(1)$ hits either zero

[7]It is simplest to apply the necessary and sufficient conditions for $P(i)$ in Caplin, Dean, Leahy (2016):

$$\sum_{\omega \in \Omega} \mu(\omega) \left\{ \frac{\exp(u(i,\omega)/\lambda)}{\sum_{j \in A} P(j) \exp(u(j,\omega)/\lambda)} \right\} \leq 1 \qquad \forall i; \tag{16}$$

with equality if $P(i) > 0$.

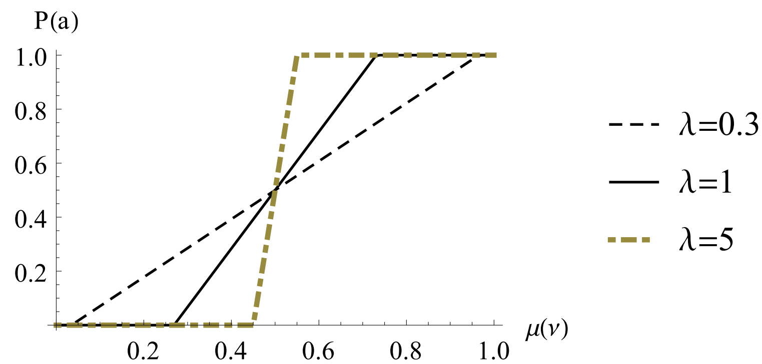

Figure 1: Effect of prior beliefs on choices, resp. effect of population shares on market shares.

or one and only the ex ante optimal option is taken.

Figure 1 presents the period zero dependence of $P(1)$ on the initial prior, $\mu(1)$, for $u(1,1) = 1$, $u(1,2) = -1$, and three different levels of λ. The higher is $\mu(1)$, the better option 1 is ex ante, the higher is $P(1)$ the unconditional probability of choosing 1. If $\mu(1)$ is sufficiently high, then the agent does not acquire any additional information, and always selects option 1. For instance, for $\lambda = 0.3$ this threshold is at $\mu(1) = 0.97$, but for $\lambda = 5$ it is already at $\mu(1) = 0.55$. On the other hand, if $P(1) \in (0,1)$, then both options are chosen in steady state, which is achieved in period one.

The figure also presents exactly what the resulting steady state market shares are by using $M(1) = P(1)$ and $g^*(1) = \mu(1)$ in (18). Notice that the slope of the dependence is greater than one, which implies that differences in realized market shares are larger than differences in shares of preference types. Learning from market shares and imperfect attention magnify the effect of popularity. For instance, if $\lambda = 5$ and if 1 is preferred by 55% or more agents, i.e., $g^*(1) \geq 0.55$, then the market share of 1 is 100%; while if 1 is preferred by 51% of agents, then its market share is 60%. The market share multiplier can be massive.

Let us now revisit the implications for individual choice. Agents choose according to the logit model with the effect of beliefs pinned down by (7). For instance, when $u(1,1) = 1$, $u(1,2) = -1$, $\lambda = 1$ and $g^*(1) = 1/2$, then while option 1 is better for all agents of type 1, only 73% of them select it, which is due to noise in signals that they acquire. Given the symmetry, $M(1) = M(2) = 1/2$. Suppose, however, $g^*(1) = 1/3$, i.e., type 1 is a minority, then $M(1) = 0.14$ and $M(2) = 0.86$. This shift in market shares biases individual choices. Using the connection to random utility models stated in Proposition 3, the choice of a

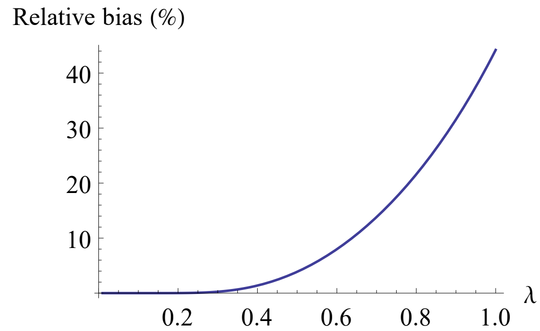

Figure 2: Bias in inference of elasticity.

rationally inattentive agent of type 1 is for $g^*(1) = 1/3$ equivalent to choice of an agent with unobserved extreme-value distributed heterogeneity with intrinsic utilities

$$U_1^1 = u(1,1) + \lambda log M(1) = u(1,1) - 1.97,$$

$$U_2^1 = u(2,1) + \lambda log M(2) = u(2,1) - 0.15.$$

The choice is biased towards alternative 2, which has higher market share. This is the effect of learning from market share. The more type 2 agents there are the more they affect market share, and the more they bias both types towards choosing their preferred option 2. When $g^*(1) = 1/3$, only 30% of type 1 agents select the option 1, which would be better for them, while 94% of type 2 agents select their favored option.

Characterizing the effect of imperfect information on inference of preferences from data is straightforward. Proposition 3 implies that if the econometrician applies standard techniques motivated by random utility models, then the inferred utility would be biased by $\lambda \log M(i)$. It would then be inferred that popular alternatives with high market shares have higher utility for all types than they do. This also affects inference of elasticities in industrial organization applications. To illustrate, consider a simple linear indirect utility function,

$$u(1,1) = 2 - \alpha p,$$

where p is price of the good 1. The task is to infer the price elasticity, $\alpha = -\partial u(1,1)/\partial p$. The standard logit-based techniques, not accounting for the bias of $\lambda \log M(i)$, would instead estimate $-\partial \left(u(1,1) + \lambda \log M(1)\right)/\partial p$. If p increases, then the steady state $M(1)$ decreases, which lowers the bias toward good 1. Inferring elasticity using the random utility approach, the price elasticity would seem higher than it is. Figure 2 presents the relative upward bias in elasticity,

$$\frac{-\partial\Big(u(1,1)+\lambda\log M(1)\Big)/\partial p}{-\partial u(1,1)/\partial p}-1,$$

as a function of λ in the neighborhood of $p=1$ and for the true elasticity $\alpha=1$. For low cost of information λ, the bias is close to 0, but it increases with λ.

6 Welfare and Policy

Our informational interpretation of discrete choice has implications for welfare and policy. In this section we discuss two issues in particular: the ability to aggregate preferences, which provides a metric for success, and the interpretation of unchosen acts, which in our model are an important source of market inefficiency.

In our model there are agents of different types who often choose options that they might prefer not to take if they had more information. This would normally complicate welfare calculations (see Bernheim and Rangel (2009)). In the standard random utility model, each agent has a different utility funciton so one also needs to make strong assumptions to make social welfare judgements. Our agents, however, solve a well-defined maximization problem (2) that is common across types. The corresponding value function $V(A,\mu)$ therefore provides a universal measure of subjective wellbeing, and we can therefore analyze the perceived effect of any policy by studying the response of $V(A,\mu)$.

There is another important notion of welfare. Since our agents may hold incorrect beliefs, a social planner who knew the true population distribution would instead weigh utilities by g^* and not by the subjective beliefs, and would want to calculate,

$$\begin{aligned}\tilde{V}(A,\mu,g^*) \quad = \quad & \max_{\{P(\omega,i)\}_{i\in A,\omega\in\Omega}} \sum_{\omega\in\Omega} g^*(\omega)\left(\sum_{i\in A} P(i|\omega)u(i,\omega)\right) \\ & -\lambda\left[\sum_{\omega\in\Omega}\mu(\omega)\left(\sum_{i\in A}P(i|\omega)\ln P(i|\omega)\right)-\sum_{i\in A}P(i)\ln P(i)\right]\end{aligned}$$

Note that here we retain μ in the information cost since the cost is driven by subjective beliefs, and the reduction in their uncertainty.

The following states that in steady state individual choice among the chosen set of alternatives is socially optimal.

Proposition 7. *In steady state, given a set of chosen options $\bar{A}$ and beliefs μ, individual choices maximize $\tilde{V}(\bar{A}, \mu, g)$.*

The proposition follows immediately from our "as if" result, Proposition 2. Since in steady state choice is made as if the prior were g^*, $V(\bar{A}, \mu) = V(\bar{A}, g^*)$, and $V(\bar{A}, g^*) = \tilde{V}(A, g^*, g^*)$. Individual choices therefore maximize both V and $\tilde{V}$ on the observed steady state choice set $\bar{A}$ and the resulting market shares are socially optimal.

There is an important caveat to Proposition 7. The proposition states that individual choice is socially optimal conditional on the set of chosen acts $\bar{A}$, but says nothing about whether the set of chosen acts is itself optimal. There is a strategic complementarity in our model: if an action is unlikely to be good, it may never be chosen. It is therefore possible that an unchosen action is superior to any chosen action, but remains unchosen because no one else is choosing it. This contrasts with the random utility model in which unchosen actions are uniformly bad: an action is not taken only if no agent finds it in their own interest to take the action.

The possibility that some good options may not be selected at all has implications for policy. Since market shares maximize welfare given the set of chosen options $\bar{A}$, it follows directly that an expansion of the set of chosen options weakly increases steady state social welfare. Given the information externality, it is possible that such an expansion improves social welfare.

Corollary 1. *Let $\bar{A}, \bar{B}$ be two sets of chosen options such that $\bar{A} \subset \bar{B}$. Then*

$$V(\bar{A}, g^*) \leq V(\bar{B}, g^*).$$

Our model provides a rationale for policies that address the nature and quality of chosen options. Traditional antitrust policy, for example, focuses on the level of prices. Competition is encouraged because it keeps prices low. In our model, policies that promote competition have additional benefits. Limiting the market shares of the market leaders provides room for new entrants whose quality would be tested by heterogeneous buyers. Selection decisions of agents testing these new entrants would generate positive information externalities for future generations of agents, which might outweigh the costs of the intervention.

7 Conclusion

The overall goal of this paper is to bring the literature on rational inattention and information acquisition closer to applied work. Key applications relate to discrete choice. While the current formulation of these models involves perfect information, our model introduces social learning and imperfect information. With regard to inference, the model allows preferences or skills to be inferred in the same way that some classical models with perfect information do, i.e., models with unobserved heterogeneity or self-selection.

Specifically, we study the evolution of market share when agents freely observe past shares and also engage in costly private information acquisition. Our analysis of steady state behavior in particular opens the doors to analysis of market behavior, policy, and to issues of inference from suitably rich data. Despite its generality, the choice behavior that emerges is very simple and strongly connected with the existing applied literature. The choice behavior follows an augmented logit model with biases given by prior knowledge. However, in steady state these biases are exactly pinned down by market shares.

These results are useful not only for their simplicity, but also for the close analytical connection to many models in the empirical literature on discrete choice, which heavily builds on logit models. However, our model implies richer choice behavior than the basic logit model. For example, independence of irrelevant alternatives is not required, and the extent of the violation is pinned down by the observable market shares.

References

[1] Ackerberg, Daniel (2003), "Advertising, Learning, and Consumer Choice in Experience Goods Markets: A structural empirical examination," *International Economic Review* 44, 1007–1040.

[2] Bartels, Larry (1996), "Uninformed Votes: Information Effects in Presidential Elections," *American Journal of Political Science* 40, 194-203.

[3] Bartoš, V., Bauer, M., Chytilová, J., and Matějka, F. (2016), "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition," *American Economic Review*, June, forthcoming.

[4] Becker, Gary (1991),"A Note on Restaurant Pricing and other Examples of Social Influences on Price," *Journal of Political Economy* 99, 1109-1116.

[5] Berry, Steven, James Levinsohn and Ariel Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica* 63, 841-890.

[6] Block, Henry David, and Jacob Marschak. (1960), "Random orderings and sto- chastic theories of responses," *Contributions to Probability and Statistics 2:* 97-132.

[7] Cai, Hongbin, Yuyu Chen, and Hanming Fang (2009), "Observational Learning: Evidence from a Randomized Natural Field Experiment," *American Economic Review* 99.3 (2009): 864-82.

[8] Caminal, Ramon, and Xavier Vives (1996), "Why Market Shares Matter: An Information Based Theory," *Rand Journal of Economics* 27, 221-239.

[9] Caplin, Andrew, Mark Dean and John Leahy (2016), "Rational Inattention, Optimal Consideration Sets, and Stochastic Choice," NYU working paper.

[10] Cardell, N. Scott (1997), "Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity," *Econometric Theory* 13(02):185-213.

[11] Cover, Thomas and Joy Thomas (2006), *Elements of Information Theory, Second Edition*, Hoboken, NJ: John Wiley & Sons.

[12] Csiszár, I. (2008), "Axiomatic characterizations of information measures," *Entropy*, 10(3), 261-273.

[13] Dasgupta, Kunal, and Jordi Mondria (2012), "Quality Uncertainty and Intermediation in International Trade," working paper.

[14] Debreu, Gerard (1960), "Review of Individual Choice Behavior by R. D. Luce," *American Economic Review* 50(1).

[15] Delli Carpini, Michael X, and Scott Keeter (1996), *What Americans Know about Politics and Why It Matters*, New Haven, CT: Yale University Press.

[16] Duflo, E. and E. Saez (2003), "The role of information and social interactions in retirement plan decisions: evidence from a randomized experiment," *Quarterly Journal of Economics* 118(3), 815-842.

[17] Erdem, T. and Michael Keane (1996), "Decision-Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science* 15, 1–20.

[18] Heckman, James J., and Bo E. Honore (1990), "The Empirical Content of the Roy model," *Econometrica* 1121-1149.

[19] Luo, Yulei (2008), "Consumption dynamics under information processing constraints," *Review of Economic Dynamics*, 11(2):366-385.

[20] Maćkowiak, Bartosz and Mirko Wiederholt (2009), "Optimal sticky prices under rational inattention," *American Economic Review* 99, 769-803.

[21] Matějka, Filip (2010), "Rationally inattentive seller: Sales and discrete pricing, " CERGE-EI Working Paper No. 408.

[22] 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.

[23] McFadden, Daniel (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka (ed.), *Frontiers in Econometrics*, 105-142, Academic Press: New York, 1974.

[24] Mondria, Jordi (2010), "Portfolio choice, attention allocation, and price comovement, *Journal of Economic Theory*, 145(5):1837-1864.

[25] Mondria, Jordi, Thomas Wu, and Yi Zhang (2010), "The determinants of international investment and attention allocation: Using internet search query data, *Journal of International Economics*, 82.1 (2010): 85-95.

[26] Moretti, Enrico (2010), "Social Learning and Peer Effects in Consumption: Evidence from Movie Sales," *Review of Economic Studies* 78, 356–393.

[27] Munshi, Kaivan (2003), "Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution," *Review of Economic Studies* 73, 175-203.

[28] Osborne Matthew (2011), "Consumer Learning, Switching Costs and Heterogeneity: A Structural Examination," *Quantitative Marketing and Economics* 9, 25-70

[29] Sims, Christopher A. (1998), "Stickiness," *Carnegie-Rochester Conference Series on Public Policy* 49, 317–356.

[30] Sims, Christopher A. (2003), "Implications of Rational Inattention," *Journal of Monetary Economics* 50, 665–690.

[31] Sims, Christopher A. (2010), "Rational inattention and monetary economics," *Handbook of Monetary Economics* 3, 155-181.

[32] Sorenson, Alan (2006), "Social Learning and Health Plan Choice," *RAND Journal of Economics* 37, 929-945.

[33] Van Nieuwerburgh, Stijn, and Laura Veldkamp (2010), "Information Acquisition and Under-Eiversification," *The Review of Economic Studies* 77, 779–805.

[34] Woodford, Michael (2009), "Information-Constrained State-Dependent Pricing," *Journal of Monetary Economics* 56, 100-124.

A Recursive Learning From Market Share

Buyers new to the market do not know their type ω. The only information freely available to agents in period $t = 0$ is their common prior G, which comprises a probability measure over distributions in $\Delta(\Omega)$. So that marginal distributions of G are well defined, we will assume that G has a continuous density on this simplex. It will be useful in what follows to define $\Gamma_0 \equiv \text{supp}(G) \subseteq \Delta(\Omega)$ as the set of possible distributions. We require that $g^* \in \text{int}(\Gamma_0)$.[8] Given G and Γ_0, we can calculate time zero agents' prior beliefs over preference types as the expected distribution of types,

$$\mu_0(\omega) = \int g(\omega) dG.$$

Agents who enter the market in periods $t > 0$ can learn about their type in part by observing past market shares. Since choice probabilities depend on agents' types, realized market shares may provide information about the distribution of types in the economy. As we shall see, this form of learning from market share involves winnowing down the set Γ_0 by eliminating distributions that are inconsistent with any observed market shares.

We now describe this process for an arbitrary period t. Let $\Gamma_t \subseteq \Gamma_0$ denote all densities in the support of Γ_0 that are consistent with all market shares observed in periods prior to t. Given Γ_t, we can calculate agents' prior beliefs over preference types as the expected distribution of types conditional Γ_t,

$$\mu_t(\omega) = \frac{1}{G(\Gamma_t)} \int_{g \in \Gamma_t} g(\omega) dG.$$

The prior μ_t determines the type dependent choice probabilities $P_t(i|\omega)$. These choice probabilities, along with the true density of types g^*, generate the realized market shares $M_t(i)$. We assume that learning is conditionally independent across agents, so that:

$$M_t(i) = \sum_{\omega \in \Omega} g^*(\omega) P_t(i|\omega). \tag{19}$$

Agents born in period $t + 1$ observe period t aggregate market shares and eliminate distributions that are inconsistent with observed market shares.[9] Γ_{t+1} includes all densities

[8] supp(G) is the set of $g \in \Delta(\Omega)$ such that every open neighborhood of g has positive measure. With $g^* \in \text{int}(\Gamma_0)$, the continuity of G ensures that the density of G is strictly positive at g^*.

[9] One feature that differentiates our approach from other models of learning from market share such as Smallwood and Conlisk (1979) is that our agents do not naively treat market share as the prior over acts

in Γ_t that generate $M_t(i)$,

$$\Gamma_{t+1} = \left\{ g \in \Gamma_t \left| \sum_{\omega \in \Omega} g(\omega) P_t(i|\omega) = M_t(i) \right. \right\}.$$

Note that if $g^* \in \Gamma_t$ then $g^* \in \Gamma_{t+1}$ as it trivially satisfies this condition. Given Γ_{t+1} period $t+1$ proceeds in a manner similar to period t completing the recursion.

B Proofs.

Proof of Lemma 1:

There are a finite number of types $\omega \in \Omega$. Hence each g may be represented by a point in the $|\Omega| - 1$ dimensional simplex in $R^{|\Omega|}$. Γ_0 is a subset of this simplex. Hence Γ_0 is the subset of an $|\Omega| - 1$ dimensional hyperplane in $R^{|\Omega|}$. This plane is the plane through g^* that is orthogonal to the unit vector

$$(g - g^*) \cdot 1 = 0$$

Define E^0 as the subspace generated by the unit vector.

Consider period t, with Γ_t and E^t. Choice in period t gives rise to a set of type specific choice probabilities $P(i|\omega)$. Let for each i, $Z^i \in R^{|\Omega|}$ denote the vector with $Z^i_\omega = P(i|\omega)$. Now the orthogonality conditions can be written as

$$(g - g^*) \cdot Z^i = 0 \qquad \forall i \text{ such that } P(\omega, i) > 0.$$

Each orthogonality condition defines a $|\Omega| - 1$ dimensional hyperplance $R^{|\Omega|}$.

There are two possibilities in period t. First, all the Z^i lie in E^t. In this case there are no new restrictions placed on the set of possible distributions, $\Gamma_{t+1} = \Gamma_t$. Learning stops and the the market has converged. Alternatively, there exists $Z^i \notin E^t$. E^{t+1} is now the space generated by E^t and the $Z^i \notin E^t$. The dimensionality of E^{t+1} is strictly greater than E^t. Γ_{t+1} is the subset of Γ_t that is orthogonal to all vectors in E^{t+1}. The dimension of Γ_{t+1} is therefore strictly less than that of Γ_t. Note, by construction, $g^* \in \Gamma_{t+1}$ if $g^* \in \Gamma_t$.

As $|\Omega|$ is finite Γ_t converges in a finite number of periods.■

Proof of Proposition 1:

but use market share to construct the prior over types.

Let us first define the following mapping $f : \Delta(\Omega) \to \mathbb{R}$:

$$f(g,i,t) = \sum_{\omega \in \Omega} \frac{g(\omega)\exp(u(i,\omega)/\lambda)}{\sum_{j \in A} P_t(j)\exp(u(j,\omega)/\lambda)}, \tag{20}$$

where P_t solves (2) for $\mu = \mu_t$. Equations (5) and (4) imply that for all i such that $M_t(i) > 0$, then $P_t(i)$ is also positive and the following holds:

$$f(g^*,i,t) = \frac{M_t(i)}{P_t(i)}. \tag{21}$$

Similarly, (5) together with the fact that the sum of probabilities in a distribution equals 1 implies:

$$f(\mu_t,i,t) = 1. \tag{22}$$

The agent in period $t+1$ knows M^t as well as P^t, and thus the agent does not deem possible those population distributions g that do not satisfy $f(g,i,t) = M^t(i)/P^t(i)$ for some i such that $M^t(i) > 0$,

$$\Gamma_{t+1} = \{g \in \Gamma_t; f(g,i,t) = M_t(i)/P_t(i)\}.$$

Now, if $f(g,i,t) = 1$ for all $g \in \Gamma_t$ and all i such that $M_t(i) > 0$, then $g^* \in \Gamma_t$ implies $f(g^*,i,t) = 1$ and thus also $M_t(i) = P_t(i)$. In this case $f(g,i,t) = M_t(i)/P_t(i)$ for all $g \in \Gamma_t$ so that $\Gamma_{t+1} = \Gamma_t$ and we have converged to a steady state $\bar{\Gamma}$.

If, on the other hand, there exist $g \in \Gamma_t$ and i such that $M_t(i) > 0$ for which $f(g,i,t) \neq 1$, then since $f(\mu_t,i,t) = 1$, $f(g,i,t)$ is linear in g and μ_t is the population distribution conditional on Γ_t, then there must exist $g' \in \Gamma_t$ for which $f(g',i,t) \neq M_t(i)/P_t(i)$ whatever $M_t(i)/P_t(i)$ is. Such g' then does not belong to Γ_{t+1}. Hence $\Gamma_{t+1} \subset \Gamma_t$.

The set of possible population distributions thus shrinks in every period, or reaches a steady state, where $M(i) = P(i)$. Γ therefore converges point-wise in $\Delta(\Omega)$. This together with (7) then immediately imply (7), too. ■

Proof of Proposition 2: In steady state, $f(g,i,t) = 1$ for all $g \in \bar{\Gamma}$ and all i such that $M(i) > 0$. Let $\bar{A} = \{i \in A | M(i) > 0\}$. Since $g^* \in \bar{\Gamma}$, $f(g^*,i,t) = 1$ for $i \in \bar{A}$. But then

$$\sum_{\omega \in \Omega} \frac{g^*(\omega)\exp(u(i,\omega)/\lambda)}{\sum_{j \in A} P(j)\exp(u(\omega,j)/\lambda)} = 1, \qquad \forall i \in \bar{A},$$

which means that P satisfies the necessary and sufficient conditions for optimality for the prior equal to g^* and an option set $\bar{A}$. ■

Proof of Proposition 5: We first fix s_1 and λ, and show that generically N combinations of $\{w_1, w_2\}$ are sufficient to $g^*(s_1, s_2)$ where N is the number of values of s_2. We then show that addition data allows the identification of λ.

Let $s_1 = \bar{s}$ and $\lambda = \bar{\lambda}$. Let s_j for $j = 1, \ldots N$, denote the possible values of s_2 conditional on $s_1 = \bar{s}$ where N is the number of possible values of s_2. Choose N distinct values for $\{w_1, w_2\}$. Index each pair by $i = 1, \ldots N$. Let γ denote the $N \times 1$ column vector with the ith row equal the observed fraction of agents who earn $\bar{w}\bar{s}$, $g_i(\bar{s})$, when the wage vector is $\{w_1, w_2\}_i$. The different wages (together with the assumed value for λ) give rise to N realizations of $P(k|s_1, s_j)$. Let $\mathbf{P}$ denote the $N \times N$ matrix whose (i, j)th element is the $P(k|\bar{s}, s_j)$ associated with $\{w_1, w_2\}_i$. Finally, define γ^* as the $N \times 1$ column vector with jth row equal to $g^*(\bar{s}, s_j)$. These definitions and (5) deliver the following system of linear equations:

$$\gamma = \mathbf{P}\gamma^*,$$

A sufficient condition for the identification of the $g^*(\bar{s}, s_j)$ is that $\mathbf{P}$ is has full rank.

We now argue that $\mathbf{P}$ is generically full rank. Let X denote the space of parameters. These include the true population distribution, the N wage pairs, and the N values of s_j. Let $\hat{X}$ denote the subspace of X in which the $P(k|s_1, s_2)$ are strictly positive. Note that $\hat{X}$ is open.

The necessary and sufficient conditions (16)

$$\sum_{\omega \in \Omega} g^*(\omega) \left\{ \frac{\exp(u(i, \omega)/\lambda)}{\sum_{j \in A} P(j) \exp(u(\omega, j)/\lambda)} \right\} = 1. \qquad \forall i$$

define the $P(j)$ as an implicit function of the parameters $x \in \hat{X}$. Note that the right-hand side is analytic as polynomials and exponentials are analytic functions and sums, products, and non-zero inverses of analytic functions are analytic. The analytic implicit function theorem then implies that the implicit function $P(x)$ is analytic as well.

Now suppose that $\mathbf{P}$ is generically not full rank. It follows that there exists some open set U in X such that $\det(\mathbf{P}) = \mathbf{0}$. But $\det(\mathbf{P}) = \mathbf{0}$ is an analytic function of $P(x)$ and $x \in \hat{X}$ and hence an analytic function of $x \in \hat{X}$, so if $\det(\mathbf{P}) = \mathbf{0}$ on U then $\det(\mathbf{P}) = \mathbf{0}$ on all of $\hat{X}$. It is easy, however, to find points in $\hat{X}$ such that P has full rank. This contradiction establishes the first part of the result.

To identify λ note that similar arguments imply that $\mathbf{P}$ and $g^*(\bar{s}, s_j)$ are analytic functions of λ. Given an additional observation of $\{w_1, w_2\}$ and the $g(\bar{s})$ associated with $\{w_1, w_2\}$ there can be only a finite number of λ consistent with this observation. If not then all λ must be consistent with this observation. In particular, as λ approaches infinity, $\sum_j P(k=1|\bar{s}, s_j)\, g^*(\bar{s}, s_j)$ approaches $P(k=1)\sum_j g^*(\bar{s}, s_j)$ where $P(k=1)$ is the observed unconditional probability of choosing $k=1$. ■

Proof of Proposition 6:

According to (5)

$$P(i|\omega) = \frac{P(i)\exp(u(i,\omega)/\lambda)}{\sum_{j\in A} P(j)\exp(u(\omega,j)/\lambda)}$$

It follows that given, $P(i), P(j) > 0$

$$\frac{P(i|\omega)}{P(j|\omega)} = \frac{P(i)\exp(u(i,\omega)/\lambda)}{P(j)\exp(u(j,\omega)/\lambda)}$$

So that

$$\frac{P(i|\omega)/P(i)}{P(j|\omega)/P(j)} = \frac{\exp(u(i,\omega)/\lambda)}{\exp(u(j,\omega)/\lambda)}$$

In steady state $P(i) = M(i)$ and the result follows.■