WORKING PAPERS

N° 1571

September 2024

Estimating Choice Models with Unobserved Expectations over Attributes

Mathias Reynaert, Wenxuan Xu and Hanlin Zhao

Toulouse School of Economics

Estimating Choice Models with Unobserved Expectations over Attributes

Mathias Reynaert Wenxuan Xu Hanlin Zhao *

September 10, 2024

Abstract

When making choices, agents often must form expectations about option attributes in the choice set. The information used to form these expectations is usually unobserved by researchers. We develop a discrete choice model where agents make choices with heterogeneous information sets that are unobserved. We demonstrate that preferences can be point-identified through a finite mixture approximation of the unobserved information structure, or set-identified using knowledge from a single agent type. These approaches are compatible with both individual- and market-level data. Applications include replicating Dickstein and Morales (2018) and estimating consumer valuations for future fuel costs without assumptions on expectation formation.

Keywords: discrete choice, unobserved information, mixture model, set identification
JEL codes: C5, C8, D8

*Reynaert: Toulouse School of Economics, University of Toulouse Capitole and CEPR, Xu and Zhao: Toulouse School of Economics, University of Toulouse Capitole. We thank seminar participants at Toulouse School of Economics, Christian Bontemps, and Arnaud Maurel, for useful discussions. We acknowledge funding by the European Union (ERC, SPACETIME, grant n° 101077168). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. We acknowledge funding from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program).

1 Introduction

Many choice models incorporate agent expectations about attributes of the choice set. For example, when choosing health insurance, individuals form expectations of their future health status (Einav et al., 2010); when purchasing vehicles, they anticipate future fuel costs (Hausman, 1979); and when deciding on educational programs, they consider expected future wages (Arcidiacono et al., 2020). In most applications of such choice models, the researcher formulates the expectations agents form about future attributes and uses these formulations in the estimation of the choice model.

Whenever the researchers' formulation of the expectation differs from how the agent forms expectations, the estimation of preferences is biased. Dickstein and Morales (2018) describe the bias from misspecified informational assumptions and propose to set identify preferences by relying on the specification of a minimal information set shared by all agents. In this paper, we present two new approaches to handle unobserved information sets when estimating choice models. First, we develop a model that can flexibly account for agents' information structures. The model we propose is estimable with individual-level and market-level choice data and is point-identified with both data types. Similar to Cunha et al. (2005), we exploit the variation between different information variables observed by the researcher and potentially used by agents to make choices. Second, we develop a novel moment-inequality approach that relies on assumptions different from the minimal information set assumption and is compatible with market-level choice data.

We specify that agents' indirect utility depends on an attribute over which they form expectations using available information. For example, in automobile choice, the attribute is fuel costs, and consumers might use variables such as stated fuel economy, fuel type, and fuel price to form their expectations about fuel costs. We allow individuals to be heterogeneous in which information variables they use. When we estimate such a model, we commonly do not observe the underlying information structure of consumers. Typically, the researcher makes an assumption, such as perfect foresight, and uses the resulting prediction of the attribute in the choice model. For instance, Grigolon et al. (2018) assumes consumers form fuel cost expectations with perfect foresight about fuel economy, mileage, and fuel prices at the moment of purchase. However, whenever the researcher's assumption deviates from the actual expectations of consumers, there is a nontrivial measurement bias problem.

We start our approach by changing the assumptions about the information sets relative to Dickstein and Morales (2018). Rather than assuming a shared minimal information set across decision-makers, we assume that the researcher observes the ex-post outcome of the attribute (e.g., fuel costs, export profits, out-of-pocket costs, wages) and a set of information variables that agents can potentially use to form expectations. We transform the ex-post product attributes observed by the researcher into a conditional expectation function of a finite number of observed information variables. We then generate all possible combinations of these information variables to create a set of discrete information types, each type associated with a single conditional expectation function. This approach builds on the intuition of Bonhomme et al. (2022) to approximate unobserved continuous heterogeneity with discrete groups.

We then plug in the set of conditional expectation functions into the choice model and estimate it as a finite mixture (i.e., latent class) model, a common statistical approach in econometrics (see Compiani and Kitamura (2016) for an overview of applications). The choice model consists of a finite number of information types, each with a specific conditional expectation function about the future attribute. Whenever each of the information types present in the data-generating process is included in our finite number of types, we show that we can point-identify the preference parameters. The key identifying variation comes from the observed heterogeneity in the information variables. We approximate the unobserved agent information structure by exploiting the combination of different potential interactions of information. While previous plug-in approaches resulted in biased preference estimates, our plug-in approach is consistent because we can flexibly approximate the unobserved information structures.

In settings with market-level data, our estimation procedure is related to Berry and Jia (2010). However, our model includes information-specific covariates that share a common coefficient across different information types. This introduces an endogeneity issue requiring information-specific instruments for resolution. We find that the information variables themselves are weak instruments because they do not capture the nonlinearity through which information affects choices well. Therefore, we adopt instruments from Chamberlain (1987) that, by construction, allow us to extract the necessary information-specific variation from the data for identification. Following Reynaert and Verboven (2014), we use a two-stage approach to estimate our parameters of interest. Simulations show that our finite mixture model provides consistent estimates.

Our approach has several advantages. First, by estimating the fraction of each information type, the model is flexible enough to detect over-specification in information heterogeneity and to examine the content of agents' information sets. For instance, if the data-generating process (DGP) does not contain an information type wrongly included in the model by the researcher, the fraction of the corresponding type is estimated to be zero. Second, our estimation procedure extends standard demand estimation tools, allowing the researcher to control for attribute endogeneity (e.g., for prices) and random coefficients in preferences, including the preference for the uncertain attribute. The model also allows us to estimate the implications of counterfactual changes in the information structure relevant to evaluating information-oriented policy.

Next, we provide a novel moment inequality approach that relies on fewer assumptions while still being compatible with individual- and market-level data. In settings with prevalent latent individual-specific information, it might be a strong assumption that we can specify all potential information variables agents can use. Instead, we prove identification by bounding the observed choices by the type-specific market shares of a single information type, similar to the approach in Gandhi et al. (2013). We assume that we know the information set of a single type in the data. For example, we assume that some individuals predicting the fuel consumption of vehicles only rely on fuel price comparisons. This allows us to estimate the conditional expectation of the uncertain attributes of these individuals and their prediction error. We show that knowledge of a single information type is sufficient to set identify parameters whenever we can specify an instrument selecting observations in the data where the observed information type has extreme errors. A feasible instrument is the specification of a cut-off for extreme values of the estimated prediction errors of the observed type. Such selection at extremes is also used in D'Haultfœuille et al. (2018). Contrary to the finite mixture approach, we do not need to specify all possible information sets, nor do we need to assume shared minimal information as in Dickstein and Morales (2018). We show theoretically and in simulations that we can set-identify the utility parameters from a single information type.

We apply our approach to two empirical settings. First, we replicate the estimation of exporter choices between destination countries in Dickstein and Morales (2018). The data contains firm-level export choices. We apply our framework by using the observed ex-post profits in the data and estimate conditional expectations firms have about these profits using combinations of information variables provided in the original data. We assume that we capture every potential information type in the DGP and specify a finite mixture maximum likelihood model. We first replicate that a naive perfect foresight plug-in estimator is biased compared to the set-identified parameters from a moment inequality approach based on the shared minimal information set across exporters. However, our model with the information mixture yields parameter estimates with overlapping confidence intervals to those from the minimal-information set identification approach. Our estimation approach directly reveals which information types explain the observed export choices in the data. We confirm that all information types that receive positive weights in the estimation results use the minimal information set specified in Dickstein and Morales (2018). The most prevalent information type additionally uses the number of existing exporters to form their expectations of future export profits.

In a second application, we revisit the estimation of fuel cost valuation in Grigolon et al. (2018). As explained above, this literature makes strong assumptions about the exact expectations consumers form regarding fuel costs and plugs in a perfect foresight prediction of fuel costs in the choice model. We apply our model by using the perfect foresight prediction of the fuel costs, but, instead of plugging in this prediction directly in the choice model, we estimate a conditional expectation function for every possible combination of the underlying information variables. We then specify the model as a finite mixture of all the possible information types and estimate their weights and preferences. We find that the estimated coefficient on fuel costs in Grigolon et al. (2018) is substantially biased upwards, and our results show no evidence that the data is generated by the perfect foresight type. We find that car buyers either form expectations based on comparisons between cars or on a rough comparison between fuel types.

Our paper adds to the structural empirical literature that models choices in situations where agent information plays a role. The minimal common information set approach with set identification is expanded in a model of physicians' choices when prescribing drugs (Dickstein et al., 2024), and in migration (Porcher et al., 2024). Our paper contributes by making a different assumption on the information structure, allowing us to achieve point identification and providing a novel moment inequality approach compatible with market-level data. Other applications, such as Arcidiacono et al. (2020); Brown and Jeon (2024); Vatter (2024), exploit specific individual-level data about the information formation of decision-makers, such as survey evidence or quality ratings. We contribute by specifying a model where researchers only need to observe the realization of the uncertain attribute and the set of potential information variables agents use to form expectations. Abaluck and Compiani (2020) prove choice data alone can be sufficient to identify preferences when consumers are uncertain about attributes. By relying on more structure and observed information variables, our approach allows us to estimate both preferences and the unobserved information structure compatible with the data. Bergemann et al. (2022) studies how to make counterfactual predictions in a setting where Bayesian individuals hold latent information that is unobserved by researchers. Our approach is consistent with any mental model to form expectations as long as we can specify it as a conditional rational expectation.

Our method can be interpreted as a generalization of the approach summarized by Cunha et al. (2005) and Cunha et al. (2010), which have been used in several empirical settings, for example, Houmark et al. (2024) and Aucejo and James (2021). This approach consists of formulating how information factors enter the decision problem of agents and separating error terms (noise). Across choices, noise should be independent of the information if the researcher correctly specifies the information. Their method can be used to test several information sets against each other. Our approach differs in that we estimate the distribution of information that fits the data best. However, the intuition is similar. In a DGP with a single homogeneous information set, our procedure would put 100% of the weight on the single correct information set because putting weight on alternative information sets would not improve the criterion function. Our approach is more general in three ways. First, the procedure selects the information sets that best explain the choices from every possible information set. This automates the researcher testing a specific information set against each other. Second, our approach allows for heterogeneity in the information sets used by agents that are identical in observables through the mixture function. Third, our approach is readily applicable to multinomial choice settings, market-level data sets, and different types of agents' prediction problems as it does not require the explicit specification of the bias stemming from misspecified information.

Our paper is structured as follows. Section 2 presents the choice model. Section 3 explains estimation issues stemming from the unobserved information structure. Section 4 introduces the finite mixture approach. Section 5 introduces the moment inequality approach. Section 6 explains estimation, including the specification of the moments to identify the information type shares. Section 7 presents simulation results. Section 8 presents the two empirical applications.

2 Choice Model

There are T markets indexed by $t \in \mathcal{T} = \{1, \ldots, T\}$. In each market t there are N_t decision-makers indexed by $i \in \mathcal{N}_t = \{1, \ldots, N_t\}$. Each individual i chooses one option indexed by $j \in \mathcal{J} = \{0, 1, \ldots, J\}$ where $j = 0$ denotes the outside option. We assume that the indirect experience utility that individual i derives from option j in market t is given by

$$u_{ijt} \equiv X_{jt}\beta + \gamma g_{ijt} + \xi_{jt} + \epsilon_{ijt}, \tag{1}$$

where $X_{jt} \in \mathbb{R}^{K_1}$ is a vector of choice characteristics, $g_{ijt} \in \mathbb{R}$ is an uncertain choice attribute that can be individual-specific and its actual value is realized only after the choice is made, $\xi_{jt} \in \mathbb{R}$ is the characteristic of option j that researchers cannot observe, and $\epsilon_{ijt} \in \mathbb{R}$ is an idiosyncratic taste shock that is i.i.d. distributed across all options j for each individual i in market t and follows a Extreme Value Type-I (EVT1) distribution. The indirect utility of the outside option is normalized as $u_{i0t} = \epsilon_{i0t}$. The vector of preference parameter of interest is $\theta = (\beta', \gamma)'$.

Denote individual i's decision variable d_{ijt}. We define $d_{ijt} = 1$ if the decision-maker chooses option j in market t and $d_{ijt} = 0$ otherwise. Individual i faces uncertainty about attribute g_{ijt} when making decisions and chooses the option j that offers the highest expected utility in market t:

$$d_{ijt} \equiv \mathbb{1}\left\{\mathcal{E}[u_{ijt}|\mathcal{I}_{it}] \geq \max_{j' \in \mathcal{J}} \mathcal{E}[u_{ij't}|\mathcal{I}_{it}]\right\}, \tag{2}$$

where $\mathcal{I}_{it}$ denotes individual i's information set about market t and $\mathcal{E}[\cdot|\mathcal{I}_{it}]$ is agent i's conditional expectation operator reflecting her beliefs. We allow the information set to be individual-specific to capture the information heterogeneity in the population. We assume individuals make rational expectations and, hence, $\mathcal{E}[\mathcal{A}|\mathcal{I}_{it}] = \mathbb{E}[\mathcal{A}|\mathcal{I}_{it}]$ for any random vector $\mathcal{A}$, where $\mathbb{E}$ is the empirical expectation from the data.

We now specify the agent's information at the decision-making stage. We assume that, when making decisions, individuals observe the choice characteristics X_{jt} and ξ_{jt}, and their taste shocks ϵ_{ijt}. However, the attribute g_{ijt} is unknown to individuals at the decision-making stage. Given this setup, we specify the decision-maker's information set as

$$\mathcal{I}_{it} = (\mathcal{W}_{it}, \{X_{jt}\}_{j \in \mathcal{J}}, \{\xi_{jt}\}_{j \in \mathcal{J}}, \{\epsilon_{ijt}\}_{j \in \mathcal{J}}, \theta'), \tag{3}$$

where $\mathcal{W}_{it}$ is the set of information variables that individual i uses to form predictions for the uncertain attribute g_{ijt}. Specifically, we define the set $\mathcal{W}_{it}$ as a collection of choice-specific information variables $k_{mjt} \in \mathbb{R}$, i.e., $\mathcal{W}_{it} = \{k_{mjt}\}_{m \in \mathcal{M}_i, j \in \mathcal{J}}$, where $\mathcal{M}_i$ is the index set of information variables used by the individual i to form her expectations of g_{ijt}. Since each individual may use different information variables to predict the uncertain attribute, the set $\mathcal{M}_i$ can be individual-specific. For instance, one individual may have an index set $\mathcal{M}_i = \{1, 2\}$ and uses two information variables k_{1jt} and k_{2jt} (i.e., $\mathcal{W}_{it} = (k_{1jt}, k_{2jt})$) to predict the uncertain attribute, while another individual i' may have $\mathcal{M}_{i'} = \{3\}$ and uses k_{3jt} instead (i.e., $\mathcal{W}_{i't} = (k_{3jt})$). Note that the realized values of the information variables k_{mjt} can vary across choices j and markets t.

At the decision-making stage, individual i needs to predict the uncertain attribute g_{ijt} based on her information set $\mathcal{I}_{it}$ as $g^e_{ijt} \equiv \mathbb{E}\left[g_{ijt}|\mathcal{I}_{it}\right]$. We further assume that agents' rational expectations only depend on the observable part of the information set $\mathcal{I}_{it}$, that is, unobservable content of the information set $\{\xi_{jt}\}_{j\in\mathcal{J}}, \{\epsilon_{ijt}\}_{j\in\mathcal{J}}$ does not help the agent forecast the uncertain attribute, i.e.,

$$g^e_{ijt} \equiv \mathbb{E}\left[g_{ijt}|\mathcal{I}_{it}\right] = \mathbb{E}\left[g_{ijt}|\mathcal{W}_{it}, \{X_{jt}\}_{j\in\mathcal{J}}\right]. \tag{4}$$

This implies that agents' predictions g^e_{ijt} are not correlated with the unobserved attribute ξ_{jt}. For simplicity, we omit $\{X_{jt}\}_{j\in\mathcal{J}}$ from the conditional expectation and assume that agents only use $\mathcal{W}_{it}$ to form their predictions in the rest of the paper.

Given (1), (3), and (4), we specify agent i's expected utility from choice j in market t as

$$u^e_{ijt} \equiv \mathcal{E}[u_{ijt}|\mathcal{I}_{it}] = X_{jt}\beta + \gamma g^e_{ijt} + \xi_{jt} + \epsilon_{ijt}. \tag{5}$$

The expected utility of the outside good remains the same as the indirect utility $u^e_{i0t} = u_{i0t} = \epsilon_{i0t}$. We decompose the expected utility (5) into the sum of three components: a mean utility term $\delta_{jt} \equiv X_{jt}\beta + \xi_{jt}$ common to all decision-makers; an individual-specific utility term $\mu_{ijt} \equiv \gamma g^e_{ijt}$, and an individual error term ϵ_{ijt}.[1]

From the perspective of researchers, we observe a random sample of T markets. For each market t, we observe choice characteristics $X_t = \{X_{jt}\}_{j\in\mathcal{J}}$ but not $\xi_t = \{\xi_{jt}\}_{j\in\mathcal{J}}$. Our main challenge is that we do not observe the expected values of the uncertain attribute $g^e_t = \{g^e_{ijt}\}_{i\in\mathcal{N}_t, j\in\mathcal{J}}$ that agents use when making decisions. Instead, as we conduct analysis ex-post the decision-making process, we observe the realized values of the uncertain attribute $g_t = \{g_{ijt}\}_{i\in\mathcal{N}_t, j\in\mathcal{J}}$. We also observe a list of K_2 information variables, $\mathcal{K}_t = \{(k_{1jt}, \ldots, k_{K_2jt})\}_{j\in\mathcal{J}}$, that individuals can potentially use to form their expectations.

Regarding the outcome variables, researchers can observe individuals' choices with different data granularities: researchers may observe individual-level choices $d_t = \{d_{ijt}\}_{i\in\mathcal{N}_t, j\in\mathcal{J}}$ or market-level proportions of each option $s_{jt} = \{s_{jt}\}_{j\in\mathcal{J}}$ that aggregate individuals' choices. Our objective is to consistently estimate the preference parameters θ based on the observed data $\{d_t \text{ or } s_t, X_t, g_t, \mathcal{K}_t\}_{t\in\mathcal{T}}$ and explore the content of individuals' information sets. We close this section with two illustrative examples of data from different aggregation levels.

Example I Researchers observe the individual-level data $\{d_{ijt}, X_{jt}, g_{jt}, \mathcal{K}_t\}_{t\in\mathcal{T}}$. For instance, this corresponds to the export market in Dickstein and Morales (2018) where researchers observe each Chilean firm i' export decisions d_{ijt} to different countries j given the year-sector t. The utility function u_{ijt} specified in Equation (2) corresponds to exporter i's profit that depends on some observed attributes X_{jt} such as the distance between Chile and the destination country $dist_j$. The profit also depends on the export revenue

[1]Our approach is compatible with any other individual unobserved utility terms, such as random coefficients.

g_{ijt} that is uncertain to exporters at the decision-making stage. Each exporter i needs to form predictions about the revenue $\mathbb{E}\left[g_{ijt}|\mathcal{I}_{it}\right]$ based on their information sets $\mathcal{I}_{it}$.[2] If the individual error term ϵ_{ijt} is EVT1 distributed, we obtain the individual choice probability:

$$\Pr(d_{ijt} = 1|\mathcal{I}_{it}) = \frac{\exp\left(\delta_{jt} + \gamma g^e_{ijt}\right)}{1 + \sum_{j=1}^{J} \exp\left(\delta_{jt} + \gamma g^e_{ijt}\right)}. \tag{6}$$

Example II Researchers observe the market-level data $\{s_t, X_t, g_t, \mathcal{K}_t\}_{t\in\mathcal{T}}$. This corresponds to the regular data of demand estimation models described in Berry (1994) and Berry et al. (1995). Researchers observe the market shares of different car models s_{jt}. For example, Grigolon et al. (2018) use this type of data to estimate consumers' valuation of fuel costs. The utility function u_{ijt} specified in Equation (2) corresponds to consumer i's utility from purchasing a car j in the year-country t. The utility depends on some observed attributes X_{jt}, such as the price p_{jt} of the car. It also depends on the future fuel costs g_{ijt} that are uncertain to consumers at the decision-making stage. Each consumer i predcits the future fuel costs $\mathbb{E}\left[g_{ijt}|\mathcal{I}_{it}\right]$ based on their information sets $\mathcal{I}_{it}$. Given the EVT1 distributed errors ϵ_{ijt}, the predicted market share s_{jt} is an integral of the individual choice probabilities over the distribution of the heterogenous predictions:

$$s_{jt}(\delta_t;\theta) = \int_{g^e_{ijt}} \frac{\exp\left(\delta_{jt} + \gamma g^e_{ijt}\right)}{1 + \sum_{j=1}^{J} \exp\left(\delta_{jt} + \gamma g^e_{ijt}\right)} dF(g^e_{ijt}), \tag{7}$$

where $\delta_t = (\delta_{1t}, \ldots, \delta_{Jt})' \in \mathbb{R}^J$ denotes the mean utility vector in market t.

3 Estimation Issues from an Unobserved Information Structure

3.1 Bias in Estimates of Preference Parameters

In this section, we discuss the estimation bias that results from not correctly specifying the information agents use when making choices. We present the bias when researchers have market-level data s_{jt}. See Dickstein and Morales (2018) for a presentation with individual-level data d_{ijt}.

Let us focus on the parameter γ which measures individuals' valuation of the expected

[2] In Dickstein and Morales (2018), the observed list of information variables $\mathcal{K}_t$ corresponds to the minimal information set that every exporter uses to predict the uncertain revenue. Thus, the observed list $\mathcal{K}_t$ is a subset of each exporter's information sets, i.e., $\mathcal{K}_t \subseteq \mathcal{I}_{ijt}$.

attribute g^e_{ijt}. Our main challenge is the unobserved expectations g^e_{ijt} in the decision utility. The value of such expectations $g^e_{ijt} \equiv \mathbb{E}\left[g_{ijt}|\mathcal{I}_{it}\right]$ depends on the individual-specific information set $\mathcal{I}_{it}$ that agents use to forecast the uncertain attribute g_{ijt}. Hence, resolving the unobserved expectations requires us to answer: What information do individuals use when forming predictions? A typical assumption that researchers often make is that individuals have perfect foresight. Thereby assuming that the unobserved expectations g^e_{ijt} coincide with the ex-post realizations of g_{ijt} and circumventing the issue of unobserved expectations.

For simplicity, consider a simple demand model for homogeneous agents, i.e., $g^e_{ijt} = g^e_{jt}, g_{ijt} = g_{jt}$, and the decision-maker predicts g_{jt} up to an error $g_{jt} = g^e_{jt} + e_{jt}$. The correctly predicted market share s_{jt} in Equation (7) becomes

$$s_{jt}(\delta_t;\theta) = \frac{\exp\left(X_{jt}\beta + \gamma g^e_{jt} + \xi_{jt}\right)}{1 + \sum_{j=1}^{J} \exp\left(X_{jt}\beta + \gamma g^e_{jt} + \xi_{jt}\right)}. \tag{8}$$

Following Berry (1994), we obtain the following linear equation that relates the observed market shares with covariates:

$$\begin{aligned} \ln(s_{jt}/s_{0t}) &= X_{jt}\beta + \gamma g^e_{jt} + \xi_{jt} \\ &= X_{jt}\beta + \gamma g_{jt} - \gamma e_{jt} + \xi_{jt} \\ &= X_{jt}\beta + \gamma g_{jt} + \tilde{\xi}_{jt}, \end{aligned} \tag{9}$$

where the last line combines the decision-maker's prediction error with the product-market-specific residual: $\tilde{\xi}_{jt} = -\gamma e_{jt} + \xi_{jt}$.

Under the perfect foresight assumption, researchers would regress market shares on the realized attributes X_{jt}, g_{jt} as stated in the last line of Equation (9). However, that equation shows that the perfect foresight assumption introduces a bias through the expectational errors e_{jt}. Indeed, $g_{jt} = g^e_{jt} + e_{jt}$ implies that $\mathbb{E}[e_{jt}|g_{jt}] \neq 0$, such that g_{jt} correlates with $\tilde{\xi}_{jt}$ and biases the estimation of the parameter γ. Because $\text{Cov}(g_{jt}, e_{jt}) > 0$ there will be an upward bias (in absolute value) in the estimation of the preference parameter γ, implying an overestimated valuation of expected attributes.

Wooldridge (2010) defines this to be a regular measurement error problem that could be potentially addressed with instrumental variables (IVs). To illustrate the idea, consider the following example when the uncertain attribute is defined as a sum of three information variables $g_{jt} = k_{1jt} + k_{2jt} + k_{3jt}$, while the agent only knows the first two information variables k_{1jt} and k_{2jt} when making predictions. We assume that the information variables are not correlated with the structural error ξ_{jt}. Individuals' expectations are $g^e_{jt} = \mathbb{E}[g_{jt}|k_{1jt}, k_{2jt}] = k_{1jt} + k_{2jt} + \mathbb{E}[k_{3jt}|k_{1jt}, k_{2jt}]$, and their expectational errors are

$e_{jt} = g_{jt} - g^e_{jt} = k_{3jt} - \mathbb{E}[k_{3jt}|k_{1jt}, k_{2jt}]$. If researchers observe the information variable k_{1jt}, it could serve as an IV if it is correlated with the endogenous regressor g_{jt} but uncorrelated with the expectational errors. This requires k_{1jt} to be uncorrelated with k_{3jt}.[3]

The validity of this IV approach depends on the researcher's ability to identify relevant information variables for each individual information set, which becomes nearly equivalent to observing the information structure. Furthermore, individual informational heterogeneity leads to a non-linear model that is non-separable in the individual-level heterogeneity. Linear IV approaches are not directly applicable in this setting.

3.2 Partial Identification

In this section, we show that the market-level data $\{s_t, X_t, g_t, \mathcal{K}_t\}_{t\in\mathcal{T}}$ cannot point identify θ and discuss what restrictions are needed to achieve point and partial identification, respectively. Dickstein and Morales (2018) show that the demand model described in Section 2 with the individual-level data $\{d_t, X_t, g_t, \mathcal{K}_t\}_{t\in\mathcal{T}}$, is not point identified.

For simplicity, we assume that $\beta = 0_{K_1}$, $\xi_{jt} = 0, \forall j, t$, and the observed list of information variables $\mathcal{K}_{jt}$ contains only one random variable. The parameter of interest hence reduces to the scalar γ and the observed data reduces to $\{s_t, g_t, \mathcal{K}_t\}_{t\in\mathcal{T}}$. None of the conclusions in the proof of the proposition below depend on these simplification assumptions.

Proposition 1 *The parameter γ is partially identified in the model described in Section 2 and the observed data $\{s_t, g_t, \mathcal{K}_t\}_{t\in\mathcal{T}}$.*

The proof is in Appendix A. The intuition is straightforward. One cannot identify the parameter γ that reflects agents' preferences associated with the unobserved predictions g^e_{ijt} when the information structure is not observable. We show that under mild specification assumptions, the parameter of interest γ can be point-identified if we know the variance of the distribution of unobserved expectations. The minimal information set assumption in Dickstein and Morales (2018) is not enough for point identification since it only represents the common knowledge of the information distribution and provides no knowledge about its variance. We need to impose restrictions on how the distribution of unobserved expectations varies across individuals to identify γ. In the finite mixture approach in Section 4, we achieve point-identification by approximating the complete unobserved information structure. Alternatively, we show that limited knowledge of the extremal properties of the

[3]For related reasons, Cunha et al. (2005) require observed information factors to be uncorrelated with unobserved information factors.

unobserved information structure also allows us to identify our parameter of interest γ without specifying the entire population's information structure. This requires weaker assumptions on the unobserved information structure. However, such inference based on the extremal information can only partially identify γ. This motivates our moment inequality approach in Section 5.

4 Finite Mixture Model

We present a semi-parametric approach based on the idea of observing an exhaustive set of information variables. We propose a two-step finite mixture model to solve the unobserved expectations problem in a setting with heterogeneous information. In the first step, we construct conditional expectations based on combinations of information variables to predict the uncertain attribute g. This results in a set of possible conditional expectation functions individuals might be using to form their expectations about the attribute. Importantly, this step does not assume the existence of a minimal information set and allows for a flexible information structure. Next, we fit these predicted conditional expectations as informational types in a finite mixture model. The conditional expectation functions are observed to vary with different information variables, informing the researcher about the proportion of information types generating the observed choice patterns in the data.

4.1 Finite Mixture Approximation of the Information Structure

We propose a semi-parametric finite mixture model to approximate the distribution of unobserved individual-specific expectations $F(g^e_{ijt})$. Our key assumption that restricts the distribution of heterogeneous expectations is the following:

Assumption 1 *The set of observed information variables $\mathcal{K}_t = \{(k_{1jt}, \dots, k_{K_2jt})\}_{j \in \mathcal{J}}$ represents a set of information variables that individuals can use to form their expectations g^e_{ijt}. No other information is available to form predictions g^e_{ijt}.*

This assumption restricts the composition of individual's information set to observable information variables. In a sense, it is the opposite of the minimal information set assumption in Dickstein and Morales (2018). Instead of assuming that we know the minimal set of observed information variables used by all individuals, we assume that we know the set of all potential information variables that any individual can use to form their expectations.

Based on Assumption 1, we can collect all the combinations of information variables in the set $\mathcal{K}_t$ to construct the set of information types potentially existing in the data. Suppose that all individuals use at least one information variable, then the total number of information types, indexed by K_3, is given by $K_3 = \sum_{k=1}^{K_2} \binom{K_2}{k} = 2^{K_2} - 1$ where K_2 is the number of observed information variables in the list $\mathcal{K}_t$.[4] Each information type, indexed by $\kappa \in \{1, \ldots, K_3\}$, corresponds to a specific information set $\mathcal{W}_{\kappa t}$ that can be used by some individuals. Relating the type-specific information set to the individual-specific information set, we have $\mathcal{W}_{it} \in \bigcup_{\kappa=1}^{K_3} \mathcal{W}_{\kappa t}, \forall i$. This framework nests the perfect foresight case when there is only one information type using the ex-post value g_{ijt} as the information variable. This discretization allows us to account for individual heterogeneity in information through a finite and countable number of information types.

Given the information set $\mathcal{W}_{\kappa t}$, we can compute the expectation $g^e_{\kappa jt} = \mathbb{E}[g_{ijt}|\mathcal{W}_{\kappa t}]$ for each information type κ. We further assume that

Assumption 2 *The functional form of the conditional expectation $\mathbb{E}[g_{ijt}|\mathcal{W}_{\kappa t}]$ is correctly specified.*[5]

Assumption 3 *Denote the fraction of each information type κ in the population by ϕ_κ, the distribution of unobserved expectations $F(g^e_{ijt})$ is approximated by a finite mixture of the expectation distributions $F(g^e_{\kappa jt})$ across all information types κ:*

$$F(g^e_{ijt}) \approx \sum_{\kappa=1}^{K_3} \phi_\kappa F(g^e_{\kappa jt}),$$

where the fraction parameters satisfy $\sum_{\kappa=1}^{K_3} \phi_\kappa = 1, \phi_\kappa \in [0, 1]$.

The idea of Assumption 3 follows from the well-known result in statistics that finite mixture models can approximate any arbitrary distribution under sufficient regularity conditions (McLachlan and Peel, 2004; Ghorbanzadeh et al., 2017; T. Tin Nguyen and McLachlan, 2020). This idea has been widely used in economic applications (see Compiani and Kitamura (2016) for a survey). For instance, both Berry and Jia (2010) and Bonhomme et al. (2022) approximate the unobserved distribution of preference heterogeneity with a finite mixture while we focus on approximating the unobserved distribution of heterogenous information (and choice attributes g^e_{ijt} here).

[4]We can also allow for an information type that does not use any information variable.

[5]In practice, we estimate with various specifications of the conditional expectation and select the one that best fits the observed correlation between the predicted attribute g_{ijt} and the information variables $\mathcal{W}_{\kappa t}$, which allows us to minimize the approximation error.

Incorporating this finite mixture model into the demand system, we can rewrite the choice probability $\Pr(d_{ijt} = 1|\mathcal{I}_{it})$ in Equation (6) (when researchers have individual-level data) or the market share s_{jt} in Equation (7) (when researchers have market-level data) as a discrete sum of type-specific choice probabilities:

$$\Pr(d_{ijt} = 1|\mathcal{I}_{it}) = \sum_{\kappa=1}^{K_3} \phi_\kappa \frac{\exp(\delta_{jt} + \gamma g^e_{\kappa jt})}{1 + \sum_{j=1}^{J} \exp(\delta_{jt} + \gamma g^e_{\kappa jt})}, \tag{10}$$

and

$$s_{jt}(\delta_t, g^e_t; \Theta) = \sum_{\kappa=1}^{K_3} \phi_\kappa s_{\kappa jt}(\delta_t, g^e_{\kappa t}; \theta) = \sum_{\kappa=1}^{K_3} \phi_\kappa \frac{\exp(\delta_{jt} + \gamma g^e_{\kappa jt})}{1 + \sum_{j=1}^{J} \exp(\delta_{jt} + \gamma g^e_{\kappa jt})}, \tag{11}$$

where $s_{\kappa jt}$ is the choice probabilities of the information type κ, $\delta_t = \{\delta_{jt}\}_{j\in\mathcal{J}}$ is the vector of mean utilities, $g^e_{\kappa t} = \{g^e_{\kappa jt}\}_{j\in\mathcal{J}}$ is the vector of expected attributes for the information type κ, $g^e_t = \{g^e_{\kappa t}\}_{\kappa=1,\ldots,K_3}$ is the $J \times K_3$ matrix of expected attributes for all information types. The vector of parameters extends to $\Theta = (\theta', \phi')'$ where θ is a vector of preference parameters and $\phi = \{\phi_\kappa\}_{\kappa=1,\ldots,K_3}$ is a vector of fraction parameters characterizing the distribution of individuals' heterogeneous information. The fraction parameters ϕ capture the information heterogeneity in the distribution of unobserved expectations, which drives heterogeneity in choices. Hence, each fraction parameter ϕ_κ can also be interpreted as the probability that an individual i, choosing the option j, belongs to a specific information type κ.

The choice probabilities in Equation (10) and (11) impose semi-parametric restrictions on the shape of the distribution of unobserved expectations. Specifically, we impose restrictions on the unknown variance of the information distribution by specifying a discrete number of values of the type-specific expectations $g^e_{\kappa jt}$. However, our specification remains flexible compared to the perfect foresight assumption or a single information set approach as we allow for the presence of different information types and estimate the fractions ϕ from the impact of information heterogeneity on individuals' choices.

Example Assume that the uncertain attribute is constant across individuals. i.e., $g_{ijt} = g_{jt}$, and that g_{jt} can be decomposed into two observed information variables as $g_{jt} = k_{1jt} + k_{2jt}$. The set of observed information variables is hence $\mathcal{K}_t = \{k_{1jt}, k_{2jt}\}$, the number of information variables is $K_2 = 2$ and the number of information types is $K_3 = 2^{K_2} - 1 = 3$. Denote the three information types as $\kappa = A, B, C$. The information sets for each type are $\mathcal{W}_{At} = \{k_{1jt}\}$, $\mathcal{W}_{Bt} = \{k_{2jt}\}$, and $\mathcal{W}_{Ct} = \{k_{1jt}, k_{2jt}\}$, respectively. Each information type forms the expectations $g^e_{\kappa jt}$ as follows: $g^e_{Ajt} = k_{1jt} + \mathbb{E}[k_{2jt}|k_{1jt}]$, $g^e_{Bjt} = k_{2jt} + \mathbb{E}[k_{1jt}|k_{2jt}]$, and $g^e_{Cjt} = k_{1jt} + k_{2jt} = g_{jt}$. Thus, the information type C corresponds to the perfect foresight type. The market share in Equation (11) now writes

$$
\begin{aligned}
s_{jt}(\delta_t, g^e_{\kappa t}; \Theta) &= \phi_A \frac{\exp(\delta_{jt} + \gamma g^e_{Ajt})}{D(\delta_t, g^e_{At}; \theta)} + \phi_B \frac{\exp(\delta_{jt} + \gamma g^e_{Bjt})}{D(\delta_t, g^e_{Bt}; \theta)} + \phi_C \frac{\exp(\delta_{jt} + \gamma g^e_{Cjt})}{D(\delta_t, g^e_{Ct}; \theta)} \\
&= \phi_A \frac{\exp(\delta_{jt} + \gamma(k_{1jt} + \mathbb{E}[k_{2jt}|k_{1jt}]))}{D(\delta_t, g^e_{At}; \theta)} + \phi_B \frac{\exp(\delta_{jt} + \gamma(k_{2jt} + \mathbb{E}[k_{1jt}|k_{2jt}]))}{D(\delta_t, g^e_{Bt}; \theta)} \\
&\quad + \phi_C \frac{\exp(\delta_{jt} + \gamma g_{jt})}{D(\delta_t, g^e_{Ct}; \theta)},
\end{aligned}
$$

where the second equation plugs in the expectations with observed information variables and $D(\delta_t, g^e_{\kappa t}; \theta) = 1 + \sum_{j=1}^{J} \exp(\delta_{jt} + \gamma g^e_{\kappa jt})$ is the exponential of the inclusive value of the information type κ.

4.2 Identification

We start by establishing identification of the fraction parameters ϕ. Given the identified fraction parameters ϕ, the identification of the remaining parameters is standard. We focus on the case where researchers only have aggregate market-level data since it is more challenging to identify the distribution of idiosyncratic expectations with market-level data. The individual-level data case can be treated similarly.

4.2.1 Identification of the Information Heterogeneity ϕ

Proposition 2 *The information heterogeneity parameters ϕ are identified in market t if the following (sufficient) conditions hold:*

1. *The matrix of type-specific choice probabilities $A_t = \{s_{\kappa jt}\}_{j \in \mathcal{J};\ \kappa=1,\ldots,K_3}$ has full column rank, where K_3 is the number of information types.*
2. *The number of inside goods J is at least $K_3 - 2$.*
3. *The fraction parameters ϕ are constant across options j.*

Proof. In each market t, one can write a system of $J + 1$ equations (including the outside option) using Equation (11) as $A_t \phi = s_t$, where the vector $\phi \in \mathbb{R}^{K_3}$ and

$$
\underset{(J+1)\times K_3}{A_t} = \{s_{\kappa jt}\}_{j \in \mathcal{J};\ \kappa=1,\ldots,K_3} = \begin{bmatrix} s_{10t} & s_{20t} & \cdots & s_{K_3 0t} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1Jt} & s_{2Jt} & \cdots & s_{K_3 Jt} \end{bmatrix}, \quad \underset{(J+1)\times 1}{s_t} = \begin{bmatrix} s_{0t} \\ \vdots \\ s_{Jt} \end{bmatrix}.
$$

It is clear then that there exists a unique vector of fractions ϕ that solves $A_t\phi = s_t$ in each market t, if the square matrix $A_t' A_t$ is invertible.[6]

One sufficient condition to identify information parameters ϕ is that the matrix A_t has full column rank (with more rows than columns, i.e., $J + 1 \geq K_3$). In this case, we need at least $J = K_3 - 1$ options (excluding the outside option) to identify those fraction parameters ϕ.[7] ■

The full column rank requirement implies that variations across columns (i.e., type-specific choice probabilities $s_{\kappa jt}$) are crucial for the identification of ϕ. All columns consist of the same choice set $\mathcal{J}$ and differ only in terms of the type-specific expectations $g^e_{\kappa jt}$. Thus, the choice variation caused by the information heterogeneity identifies ϕ.

We highlight that the identification argument above holds within each market t. Thus, variations across markets are not necessary to identify the information structure ϕ. Consequently, the information parameters ϕ can be market-specific ϕ_t if there is enough variation in expectations $g^e_{\kappa jt}$ across types and options for identification. Additionally, the number of information types K_3 can also be market-specific if we observe different information variables across markets. In our empirical application, we assume constant fraction parameters ϕ across markets, hence observations from different markets improve estimation by increasing the sample size. [8]

Our framework allows for testing the content of agents' information sets. Given an information type κ, we can test whether the information variables in the set $\mathcal{W}_{\kappa t}$ are relevant for forming expectations. If the fraction parameter ϕ_κ is significantly different from zero, it implies that the information variables in the set $\mathcal{W}_{\kappa t}$ are used by some individuals to form their expectations. This test can be conducted for each information type κ.

4.2.2 Identification of the Preference θ

Once the information structure ϕ is identified, we are back to a typical demand estimation setting, and we can extend the model in Equation (11) to account for additional preference heterogeneity, e.g., by introducing random coefficients β_i and γ_i. With knowledge of

[6] One can add the summation restriction on fraction parameters $\sum_{\kappa=1}^{K_3} \phi_\kappa = 1$ into the system of equations and adjust the matrices A_t, s_t accordingly. A similar identification condition can be derived.

[7] If we consider the summation restriction on fraction parameters, $\sum_{\kappa=1}^{K_3} \phi_\kappa = 1$, exact identification can be achieved with $K_3 - 2$ options (excluding the outside option).

[8] In practice, a constant fraction assumption implies that each market contains the same proportion ϕ_κ of individuals that use a specific set of information variables (e.g., $\{k_1, k_2\}$). However, this does not imply that the same proportion of individuals has the same expectations across markets as the value of information variables $\{k_1, k_2\}$ still varies across both markets and options.

ϕ, the market share system is invertible, and thereby, identification of the parameters θ follows Berry and Haile (2014). Identifying the distribution of random coefficients relies on variation across markets, which does not rely on the information heterogeneity as long as the fraction parameters ϕ are constant across markets.

4.2.3 Discussion

Finally, we discuss disentangling preference heterogeneity from information heterogeneity in our model. The preference heterogeneity is captured by the random coefficients on the standard linear parameters β and on the valuation of expected attributes γ. It restricts the distribution of preference parameters. The objective of using random coefficients is to break the Independence of Irrelevant Alternatives (IIA) assumption of regular logit models, which does not solve any measurement error problem described in Section 3. By contrast, the information heterogeneity is captured by the fraction parameters ϕ and the plug-in expectations $g^e_{\kappa jt}$ that characterize the distribution of individuals' heterogeneous information. It restricts the distribution of the regressor g^e_{ijt} in the demand system. The objective of modeling the information heterogeneity is to approximate the unobserved information distribution $F(g^e_{ijt})$ with a finite mixture, which solves the measurement error problem.

It is often difficult to distinguish preference and information heterogeneity because one cannot attribute decisions to a strong preference for an option, or owing specific information about that choice. We propose that this identification problem can be overcome with two additional assumptions. First, we need to be able to recover the specific effect of information on choice. We do this using the rational expectation assumption and exploiting both within- and cross-market variations to capture the way decision-makers use different information to predict the uncertain attribute. Second, we need to disentangle information from preferences. Here, we assume that preferences are stable across options and markets while information variables realize differently across options and markets. This allows us to identify the distribution of information separately with the distribution of preferences. The difficulty with disentangling preferences and information is also discussed in papers that deal with unobserved choice sets. See Barseghyan et al. (2021) or Molinari (2020) for more examples.

5 Moment Inequalities with Extremal Information

The semi-parametric finite mixture model developed so far relies on the specification of a complete set of information variables that any agent might use as stated in Assumption 1. The mixture model has two limitations. First, there is a curse of dimensionality: the number of information parameters ϕ increases exponentially with the number of information variables. Second, researchers might not be able to posit an exhaustive set of information variables. The mixture model will be biased when agents have latent information not included by the researcher among the potential information variables.

We now turn to a partial identification approach to make valid inferences about preferences without specifying a comprehensive set of information variables. We contribute by deriving novel moment inequalities based on evaluating choices when the information is at extreme values. The approach is compatible with market-level data (while previous moment inequality approaches, such as those based on shared minimal information, are only compatible with individual-level data). The cost of relaxing Assumption 1 is that we cannot restrict the distribution of individuals' expectations, and many unobserved information types could explain the data. Thus, we only have partial identification as illustrated in Section 3.2.

We argue that partial knowledge about the expectation distribution, in particular its properties on the extremums, is sufficient to set identify the preference parameters θ.[9] This partial identification requires the researcher to observe at least the information set of a single information type existing in the data and for that type to be the extremal information type at some realizations of the information variables.

For any instruments Z_{jt}, we define conditional moment inequalities as

$$\mathbb{E}\left[\begin{matrix} m_{\max}(s_{jt}, g_{jt}, X_{jt}; \theta) \\ m_{\min}(s_{jt}, g_{jt}, X_{jt}; \theta) \end{matrix} \middle| Z_{jt}\right] \geq 0, \tag{12}$$

where the two moment functions are defined as

$$\begin{aligned} m_{\max}(s_{jt}, g_{jt}, X_{jt}; \theta) &= -\log \frac{s_{jt}}{s_{0t}} - \gamma \max_{\kappa} \{e_{\kappa jt}\} + \gamma g_{jt} + X_{jt}\beta + \xi_{jt}, \\ m_{\min}(s_{jt}, g_{jt}, X_{jt}; \theta) &= \log \frac{s_{jt}}{s_{0t}} + \gamma \min_{\kappa} \{e_{\kappa jt}\} - \gamma g_{jt} - X_{jt}\beta - \xi_{jt}. \end{aligned} \tag{13}$$

Let Θ denote the set of all possible parameter values θ and Θ_0 denote the subset of those values consistent with the conditional moment inequalities defined in (12), we have

Theorem 1 *Let θ^* be the true parameter defined in the model* (11). *Then $\theta^* \in \Theta_0$.*

[9] D'Haultfœuille et al. (2018) presents a similar idea to solve selection bias by focusing on extremal points.

Theorem 1 states that the extremal moment inequalities are consistent with the true parameter value θ^*. We provide an intuitive explanation of Theorem 1 below. The formal derivation is in Appendix B.

For ease of illustration, we rewrite inequalities (13) as two bounds for the observed log-ratio of market shares $\log \frac{s_{jt}}{s_{0t}}$:

$$-\gamma \min_{\kappa} \left\{e_{\kappa jt}\right\} + \gamma g_{jt} + X_{jt}\beta + \xi_{jt} \leq \log \frac{s_{jt}}{s_{0t}} \leq -\gamma \max_{\kappa} \left\{e_{\kappa jt}\right\} + \gamma g_{jt} + X_{jt}\beta + \xi_{jt}. \quad (14)$$

Inequalities (14) illustrate that the extremum ratios of choice probabilities are characterized by individuals who commit extremal errors when predicting the uncertain attribute g_{jt}. The inequalities provide an upper (lower) bound of the log-share ratio with the maximum (minimum) expectational errors that agents commit. These inequalities mirror those found in Gandhi et al. (2013) who employed linear conditional inequalities with market-level data to perform robust inference in contexts featuring zero market shares. The two sides of the inequality (14) are not redundant. Holding fixed all parameters except γ and assuming without loss of generality that $\gamma < 0$, the right-hand side of (14) is decreasing in γ and hence identifies an upper bounds on γ. The left-hand side is a lower bound on γ.

Figures 5a and 5b graphically illustrate the power of our inequalities (14) in identifying the correct values of the parameter γ. The right-hand side of (14) is plotted in blue while the left-hand side in green. The observed middle part of (14) is plotted in red. Given the true parameter value $\gamma_0 = -1.5$, Figure 5a shows that the maximal error pseudo-type in blue always bounds the market share ratio from above while the minimal error pseudo-type in green bounds the data from below. However, given a wrong parameter value $\gamma = 2 \neq \gamma_0$, our inequality is violated, which will then rule out such wrong parameter value from the identified set $\Theta_{0,\gamma}$, as depicted in Figure 5b.

Inequalities in (14) are not estimable since they contain unobserved terms ξ_{jt}, $\max_{\kappa} \left\{e_{\kappa jt}\right\}$ and $\min_{\kappa} \left\{e_{\kappa jt}\right\}$. Our objective is to express them with observed data by implementing appropriate instruments Z_{jt}. Regarding the unobserved attribute ξ_{jt}, we assume it has a zero-mean $\mathbb{E}\left[\xi_{jt}|Z_{jt}\right] = 0$ conditional on regular instruments $Z_{jt} = \{X_{jt}\}$.[10] For simplicity, we assume $\gamma < 0$ in the following discussion.

To build intuition for the selection at extremes, we start with an example that relies on computing the unobserved errors at observed extremal points to obtain values for $\mathbb{E}\left[\max_{\kappa} \left\{e_{\kappa jt}\right\} |Z_{jt}\right]$ and $\mathbb{E}\left[\min_{\kappa} \left\{e_{\kappa jt}\right\} |Z_{jt}\right]$. For the sake of the example, as-

[10]Our approach can account for cases where the observed attribute vector X_{jt} contains endogenous attributes such as prices that are correlated with the unobserved attribute ξ_{jt}. Then one needs to include extra instruments such as regular price instruments Z^p_{jt} into the instrument set $Z_{jt} = \{X_{jt}, Z^p_{jt}\}$.

sume the uncertain attribute $g_{jt} \in [a,b]$, type-specific expectations $g^e_{\kappa jt} \in [c,d]$ and that the researcher observes a,b,c,d.[11] This implies the prediction errors have bounds $e_{\kappa jt} \equiv g_{jt} - g^e_{\kappa jt} \in [a-d, b-c]$. This leads to the following identification assumption:

Assumption 4 *We assume that a,b,c,d are observed by the researcher and denote the bounds of the uncertain attribute $g_{jt} \in [a,b]$ and type-specific expectations $g^e_{\kappa jt} \in [c,d]$ so that $\max_\kappa \{e_{\kappa jt}\}, \min_\kappa \{e_{\kappa jt}\} \in [a-d, b-c]$. We also assume $\mathbb{E}\left[\xi_{jt}|Z_{jt}\right] = 0$ where $Z_{jt} = \{X_{jt}\}$.*

Assumption 4 imply the following moment functions:

$$m_{\max}(s_{jt}, g_{jt}, X_{jt}; \theta) = -\log \frac{s_{jt}}{s_{0t}} - \gamma(b-c) + \gamma g_{jt} + X_{jt}\beta,$$
$$m_{\min}(s_{jt}, g_{jt}, X_{jt}; \theta) = \log \frac{s_{jt}}{s_{0t}} + \gamma(a-d) - \gamma g_{jt} - X_{jt}\beta,$$

which are easy to implement but can lead to large identified set when the differences $a-d$ and $b-c$ are large.

The example above illustrates that the key to deriving estimable inequalities is to find tight bounds for the unobserved expectations $\mathbb{E}\left[\max_\kappa \{e_{\kappa jt}\} |Z_{jt}\right], \mathbb{E}\left[\min_\kappa \{e_{\kappa jt}\} |Z_{jt}\right]$. It is difficult to recover the full distribution of the maximum (minimum) errors $\max_\kappa \{e_{\kappa jt}\}$ ($\min_\kappa \{e_{\kappa jt}\}$) and then compute the corresponding conditional expectations. However, it is possible to observe data points where the extremal errors are dominated by one single type conditional on a selection instrument Z^S_{jt}. Focusing on those specific observations and knowing the information set of that single type allow us to recover the unobserved expectations $\mathbb{E}\left[\max_\kappa \{e_{\kappa jt}\} |Z_{jt}\right], \mathbb{E}\left[\min_\kappa \{e_{\kappa jt}\} |Z_{jt}\right]$. Taking the maximum errors for illustration, a proposal for an instrument is an indicator variable capturing when the value of one single type $\kappa_{\max}$'s error exceeds a certain threshold a i.e., $Z^S_{jt} = \{e_{\kappa_{\max} jt} \geq a\}$.[12] When individuals of type $\kappa_{\max}$ make the largest prediction error, i.e., $e_{\kappa_{\max} jt} \geq e_{\kappa jt}, \forall \kappa$, we can recover the unobserved expectations $\mathbb{E}\left[\max_\kappa \{e_{\kappa jt}\} |Z^S_{jt}\right] = \mathbb{E}\left[e_{\kappa_{\max} jt}|Z^S_{jt}\right]$.

[11]If the researcher specifies each information type correctly, i.e., there is no unobserved information variable or information type, then the assumption that the researcher observes c,d is valid. However, when there are concerns of misspecification, researchers may not observe correct values of c,d. In this latter case, we can, for instance, arbitrarily specify extreme values of c,d that are robust to potential misspecification issues. Alternatively, we can assume that agents do not make extreme predictions beyond the scope of the ex-post observations g_{jt} so that type-specific expectations are also bounded by the same values $g^e_{\kappa jt} \in [a,b]$.

[12]Recall that expectational errors are defined as $e_{\kappa jt} \equiv g_{jt} - g^e_{\kappa jt}$, where the observed ex-post attribute is a function of observed (and unobserved if there are concerns of misspecification) information variables in the set $\mathcal{K}_t$ i.e., $g_{jt} \equiv f_g(\mathcal{K}_t)$ and the type-specific prediction is a function of information variables in the subset $\mathcal{W}_{\kappa t} \subseteq \mathcal{K}_t$. i.e., $g^e_{\kappa jt} \equiv f_{g^e}(\mathcal{W}_{\kappa t})$. Thus, we can express the type-specific error as a function of information variables $e_{\kappa jt} = f_e(\mathcal{K}_t)$. Then a sufficient condition for the selection $e_{\kappa_{\max} jt} \geq a$ can be an information variable k_{jt}, dominating the type $\kappa_{\max}$'s error, exceeds a certain threshold a, i.e., $Z^S_{jt} = \{k_{jt} \geq a\}$.

This approach requires the researcher to observe the information set of at least a single information type and to be able to select observations where that information type makes extreme errors using a selection instrument Z^S_{jt}. Adding regular instruments for the unobserved attribute ξ_{jt}, the instrument set becomes $Z_{jt} = \{X_{jt}, Z^S_{jt}\}$. We formalize our approach as:

Assumption 5 *Let $\kappa_{\max}, \kappa_{\min}$ denote the two extremal information types, conditional on the observed selection instruments $Z^S_{\max jt}, Z^S_{\min jt}$ respectively. Conditional on $Z^S_{\max jt}$, the maximum type $\kappa_{\max}$ predicts the uncertain attribute g_{jt} with the largest error $e_{\kappa jt} \equiv g_{jt} - g^e_{\kappa jt}$, among all the information types κ, for option j in market t, i.e., $e_{\kappa_{\max} jt} \mid Z^S_{\max jt} = \max_{\kappa} e_{\kappa jt} \mid Z^S_{\max jt}$. The minimum type $\kappa_{\min}$ is defined similarly. We assume that*

(i) We observe at least one information set W_{κ_O} for some individuals that exist in the data.

(ii) The observed information type κ_O is the extremal type $\kappa_{\max}$ ($\kappa_{\min}$) at data points selected by the instrument $Z^S_{\max jt}$ ($Z^S_{\min jt}$).

(iii) $\mathbb{E}\left[\xi_{jt} | Z_{jt}\right] = 0$ where $Z_{jt} = \{X_{jt}, Z^S_{\max jt}, Z^S_{\min jt}\}$.

Assumptions 5 imply the following moment functions:

$$\begin{aligned} m_{\max}(s_{jt}, g_{jt}, X_{jt}; \theta) &= -\log \frac{s_{jt}}{s_{0t}} + \gamma g^e_{\kappa_{\max} jt} + X_{jt}\beta, \\ m_{\min}(s_{jt}, g_{jt}, X_{jt}; \theta) &= \log \frac{s_{jt}}{s_{0t}} - \gamma g^e_{\kappa_{\min} jt} - X_{jt}\beta, \end{aligned} \tag{15}$$

where $g^e_{\kappa_{\max} jt} = g_{jt} - e_{\kappa_{\max} jt}$ and $g^e_{\kappa_{\min} jt} = g_{jt} - e_{\kappa_{\min} jt}$.

The selection instrument implies that our inference is only based on a subsample of observations. When information variables vary strongly over markets, we expect to obtain more precise confidence regions as it allows us to select more observations where particular information variables are very different from others.

Consider fuel costs as a practical example. We observe fuel prices and fuel economy, and we might assume that when fuel prices are extreme, individuals predicting fuel costs by relying solely on fuel economy information make more pronounced errors than any other information type. We only need to assume that some information types are making a prediction based on fuel economy and that one of these information types' prediction errors is larger than any other error when fuel prices are above or below a threshold.

6 Estimation

6.1 Finite Mixture Approach

6.1.1 Estimation with Individual-level Data

Given individual-level data, we estimate our parameters of interest with the maximum likelihood. The likelihood function is

$$\mathcal{L}(\Theta|d,\mathcal{I}) = \prod_{ijt} \left[\Pr(d_{ijt}=1|\mathcal{I}_{it};\Theta)\right]^{d_{ijt}} \left[1-\Pr(d_{ijt}=1|\mathcal{I}_{it};\Theta)\right]^{1-d_{ijt}}. \quad (16)$$

where the choice probability $\Pr(d_{ijt}=1|\mathcal{I}_{it};\Theta)$ is defined in Equation (10).

6.1.2 Estimation with Market-level Data

We propose to extend the GMM approach based on Berry et al. (1995), which leverages on the assumption of mean independence of the structural errors ξ_{jt}, $\mathbb{E}\left[\xi_{jt}|Z_{jt}\right]=0$, where $Z_{jt}\in\mathbb{R}^{K_4}$ represents the vector of K_4 instruments. The estimation algorithm searches over parameter vectors Θ to minimize a criterion function based on unconditional sample moment conditions formed between the structural errors ξ_{jt} and the instruments Z_{jt}.

Given a value for the parameters ϕ, we can employ a contraction mapping to recover the mean utility terms δ_t from the observed market shares s_t. Next, we regress the mean utility δ_t on the product characteristics X_t (including endogenous prices p_t) to obtain the residuals. These residuals serve as estimates of the structural errors $\xi_t(\Theta)=\delta_t(s_t-X_t\beta)$, which enables us to compute the GMM objective function that we minimize with constraints:

$$\min_{\Theta} \xi(\Theta)'Z'WZ\xi(\Theta)$$

$$\text{subject to } \sum_{\kappa=1}^{K_3}\phi_\kappa = 1, \phi_\kappa\in[0,1],$$

where the vectors and matrices are stacked over all markets t, W is a $K_4\times K_4$ squared weighting matrix.

We assume that the K_1 observed product characteristics X_{jt} and K_2 information variables k_{mjt} are mean independent of the structural error. Our identification condition writes

$$\mathbb{E}\left[\xi_{jt}|X_{jt},\{k_{mjt}\}_{m=1}^{K_2}\right]=0.^{13} \quad (17)$$

[13]When there are endogenous regressors, such as prices, among the observed regressors X_{jt}, we can use some observed exogenous cost shifters and calculate the BLP instruments and differentiation IVs, denoted as Z^P, to solve the endogeneity issue. The identification condition rewrites $\mathbb{E}\left[\xi_{jt}|X_{jt},\{k_{mjt}\}_{m=1}^{K_2},Z_{jt}^P\right]=0$.

However, we argue that regular information variables are not strong instruments to identify the preference parameter γ and information parameter ϕ in our model. The reasons are twofold. First, identification of the information structure ϕ relies on the variations of the market shares across information types, i.e., the impact of plug-in expectations $g^e_{\kappa jt}$ on choices s_{jt}. Consequently, it is not the information variable per se but those type-specific expectations that provide exogenous variations for identification. The expectation $g^e_{\kappa jt}$ is a non-linear function of the information variable k_{mjt} and our model in Equation (11) is also non-linear in the expectation $g^e_{\kappa jt}$.[14] Therefore, we need a non-linear IV to pick out the right variation in this conditional expectation. Second, our model in Equation (11) is specified with all the information types implied by the Assumption 1 while some of the plug-in types may not exist in the data. In this case, there is no reason to believe that the irrelevant information variables satisfy the mean independence assumption, which creates a weak IV issue. Both reasons suggest that we need to construct non-linear instruments that provide exogenous variations reflecting the impact of expectations on choices and are robust to over-saturated irrelevant information types.

We find that in applications Chamberlain's (optimal) instruments (Chamberlain, 1987) work well to address the non-linearities through its statistical form as derivatives targeting each parameter. These instruments correspond to the expected Jacobian matrix $\mathbb{E}\left[\frac{\partial \xi_{jt}(\Theta)}{\partial \Theta'}\middle| Z_{jt}\right]$ of the structural error ξ with respect to the parameter vector Θ, conditional on exogenous variables Z_{jt}. First, it constructs the required exogenous variations via a non-linear operator (i.e., the conditional expectation) on plain instruments Z_{jt} such as information variables. Second, as first derivatives, they are tailored to each parameter ϕ and γ. If there is an irrelevant information type κ, then one can expect that the derivative of the structural error with respect to that irrelevant parameter ϕ_κ is asymptotically zero. In other words, Chamberlain's instruments select out irrelevant types based on data by assigning negligible weights to moments associated with non-existing types while favoring other regular moments. This helps smooth the objective function and ensures consistent estimates. To summarize, plug-in expectations combined with the finite mixture approximation achieve the identification of the parameters γ and ϕ, while Chamberlain's instruments serve as strong instruments to consistently estimate parameters.

Reynaert and Verboven (2014) highlight the performance of Chamberlain's instruments in identifying nonlinear parameters and provide a procedure to compute an empirical

[14]The unknown expectation form $\mathbb{E}$ matters to construct a strong IV. In practice, we plot the joint and conditional distribution between the ex-post attribute and information variables to acquire knowledge about the potential shape of those conditional expectations $g^e_{\kappa jt}$. Then we experiment with different forms of expectations such as linear, kernel, polynomial of degree two and random forest models.

approximation of these instruments. Following their work, our estimation involves a two-stage process. We first estimate the model using regular instruments $(X_{jt}, \{k_{mjt}\}_{m=1}^{K_2})$ in the first stage. Then, we compute Chamberlain's instruments as the expected value of the derivatives of the structural error ξ_t with respect to the parameters $(\phi', \gamma)'$ conditional on the first stage's regular instruments. These derivatives are evaluated at the first-stage estimates and mean utility terms. Finally, we estimate the model again in a second stage leveraging Chamberlain's instruments.

6.2 Moment Inequalities Approach

The estimation procedure consists of three steps. First, we specify the information set $\mathcal{W}_{\kappa_O t}$ for the observed extremal information type κ_O, estimate her predictions $g^e_{\kappa_O jt} = \mathbb{E}[g_{jt}|\mathcal{W}_{\kappa_O t}]$ and the associated prediction errors $e_{\kappa_O jt}$. Second, we form the instrument Z^S_{jt} that selects observations where the prediction errors $e_{\kappa_O jt}$ reach their extreme values. Finally, we adapt the general conditional moment inequality framework of Andrews and Shi (2013) to estimate the confidence region for our parameter of interest based on the conditional moment inequalities defined in (15).[15]

7 Simulations

We conduct simulations to illustrate the performance of our finite mixture approach and the moment-inequalities with market-level data. The simulations highlight the importance of Chamberlain's instruments and showcase that we can find an informative set with the moment inequality approach.

7.1 Setup

Our simulation is based on the example described in Section 4.1 with three information types. We simulate 1000 datasets, each consisting of $T = 25$ markets and $J = 10$ products. The indirect utility is defined as

$$u_{ijt} = \beta_0 + \beta_x x_{jt} - \alpha p_{jt} + \gamma g_{jt} + \xi_{jt} + \epsilon_{ijt}.$$

[15]Recent work in Andrews et al. (2023) and Cox and Shi (2023) have expanded the set of inference procedures that can be done in settings with linear conditional moment inequalities, but applying Andrews and Shi (2013) have been sufficient to obtain reasonable results in our simulations.

The product characteristic is uniformly distributed $x_{jt} \overset{\text{iid}}{\sim} U(1.5, 2.5)$. We simulate the price $p_{jt} = 1 + z_{jt} + \nu_{jt}$ to be exogenous where the cost shifter $z_{jt} \overset{\text{iid}}{\sim} U(0,1)$ and the cost shock $\nu_{jt} \overset{\text{iid}}{\sim} U(-0.25, 0.25)$. The uncertain attribute is assumed to be constant across individuals $g_{ijt} = g_{jt}$ and equals the sum of two observed information variables $g_{jt} = k_{1jt} + k_{2jt}$. The two information variables are independent of each other. The first information variable is uniformly distributed $k_{1jt} \overset{\text{iid}}{\sim} U(0,1)$ with expectation $\mathbb{E}[k_{1jt}] = 0.5$ and variance $\text{Var}[k_{1jt}] = 1/12$. The second follows a log-normal distribution $\log(k_{2jt}) \overset{\text{iid}}{\sim} N(0,1)$ with expectation $\mathbb{E}[k_{2jt}] = \exp(0.5)$ and variance $\text{Var}[k_{2jt}] = e(e-1)$ where e is Euler's number. The demand shifters are uniformly distributed $\xi_{jt} \overset{\text{iid}}{\sim} U(-1,1)$. The demand shocks follow an EVT1 distribution. The preference parameters are $\theta = (\beta_0, \beta_x, \alpha, \gamma)' = (1, 1, 1.5, -1.5)'$.

There are three information types in the simulations, indexed by $\kappa = A, B, C$, where the information set of each type is $\mathcal{W}_{At} = \{k_{1jt}\}$, $\mathcal{W}_{Bt} = \{k_{2jt}\}$, and $\mathcal{W}_{Ct} = \{k_{1jt}, k_{2jt}\}$. The proportion of each information type is $\phi_A = 0.15, \phi_B = 0.5, \phi_C = 0.35$. Given this information structure, the average market share of the outside option is around 21.7%. Each information type forms the expectations $g^e_{\kappa jt}$ as follows: $g^e_{Ajt} = k_{1jt} + \mathbb{E}[k_{2jt}|k_{1jt}]$, $g^e_{Bjt} = k_{2jt} + \mathbb{E}[k_{1jt}|k_{2jt}]$, and $g^e_{Cjt} = k_{1jt} + k_{2jt} = g_{jt}$. Here, the information type C corresponds to a perfect foresight type. The expectations $g^e_{\kappa jt}$, the expectational errors $e_{\kappa jt}$, and the variances of the expectational errors $\text{Var}(e_{\kappa jt})$ are as follows:

Types	Expectations	Expectational errors	Error variances
A	$g^e_{Ajt} = k_{1jt} + \exp(0.5)$	$e_{Ajt} = k_{2jt} - \exp(0.5)$	$e(e-1)$
B	$g^e_{Bjt} = k_{2jt} + 0.5$	$e_{Bjt} = k_{1jt} - 0.5$	$1/12$
C	$g^e_{Cjt} = k_{1jt} + k_{2jt} = g_{jt}$	$e_{Cjt} = 0$	0

Furthermore, we extend this dataset of three information types to allow preference heterogeneity with random coefficients on the valuation of expected attributes γ. The random coefficient is specified as $\gamma_i = \gamma + \sigma_\gamma v_i$ where $\gamma = -1.5$ and $\sigma_\gamma = 0.5$. The distribution of γ_i is constructed by 500 draws of v_i from the standard normal distribution and we have $\gamma_i \overset{\text{iid}}{\sim} N(-1.5, 0.25)$. To ensure that this setup generates a similar average market share of the outside option as the previous data, we adjust the distribution of the exogenous attribute $x_{jt} \overset{\text{iid}}{\sim} U(2,3)$. The average market share of the outside option is around 19.42% in such datasets.

Finally, we simulate a third data that only contains two information types $\kappa = B, C$ with fractions $\phi_B = 0.4, \phi_C = 0.6$ to test if our method can recover consistent estimates even when there are fewer information types in the data than specified. The average market share of the outside option is around 17.57% in such datasets.

Table 1: Demand Estimation in the Three-type Monte Carlo Simulation

		Three-Type						Three-type & RC		
		Regular IV			Chamberlain's IV			Chamberlain's IV		
param.	true	est.	st. err.	bias	est.	st. err.	bias	est.	st. err.	bias
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
β_0	1	0.88	(0.228)	0.12	1.00	(0.023)	0.00	0.97	(0.435)	0.03
β_x	1	1.01	(0.021)	-0.01	1.01	(0.006)	-0.01	1.00	(0.133)	0.00
α	1.5	1.51	(0.016)	-0.01	1.50	(0.006)	0.00	1.50	(0.123)	0.00
ϕ_A	0.15	0.14	(0.021)	0.01	0.15	(0.003)	0.00	0.15	(0.024)	0.00
ϕ_B	0.5	0.46	(0.089)	0.04	0.50	(0.008)	0.00	0.48	(0.093)	0.02
ϕ_C	0.35	0.40	(0.028)	-0.05	0.35	(0.104)	0.00	0.37	(0.099)	-0.02
γ	-1.5	-1.42	(0.176)	-0.08	-1.51	(0.009)	0.01	-1.49	(0.115)	-0.01
σ_γ	0.5							0.48	(0.124)	0.02

7.2 Simulation of Finite Mixture Approach

We estimate the model without preference heterogeneity using the GMM two-step procedure described in Section 6. Specifically, we employ regular instruments in the first stage and calculate the approximated Chamberlain's instruments for parameters γ and ϕ based on the first stage estimates following Reynaert and Verboven (2014).[16] Then, we replace the information variables k_{1jt}, k_{2jt} in the set of regular instruments with Chamberlain's instruments in the second stage.

When estimating the dataset with preference heterogeneity, it is crucial to include strong instruments to obtain consistent estimates. We find that the first-stage estimates are usually heavily biased due to the lack of strong instruments, affecting the quality of the Chamberlain's instruments in the second stage. To address this issue, we implement a three-step procedure by constructing Chamberlain's instrument twice. We calculate a second-stage Chamberlain's instruments based on the first-stage estimates and use those to construct a third-stage Chamberlain's instruments for estimation in the final stage.

Table 1 displays the estimation results in one simulated data where all three information types are present in the DGP. Columns (3)-(8) show that the bias of estimates using Chamberlain's instruments is smaller than those using regular instruments. Especially for the intercept, information parameters ϕ and valuation of the expected attribute γ, the bias is significantly reduced when using Chamberlain's instruments. This justifies the Chamberlain's instruments as strong instruments. As expected, the standard errors are

[16]Specifically, regular instruments in our simulation are (constant, $x_{jt}, p_{jt}, k_{1jt}, k_{2jt}$).

Table 2: Demand Estimation in the Two-type (B, C included) Simulation

		regular IV			Chamberlain's IV		
param.	true	est.	st. err.	bias	est.	st. err.	bias
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_0	1	4.02	(0.243)	-3.02	1.06	(0.168)	-0.06
β_x	1	0.92	(0.183)	0.08	1.01	(0.084)	-0.01
α	1.5	1.57	(0.139)	-0.07	1.51	(0.083)	-0.01
ϕ_A	0	0.01	(0.011)	-0.01	3.30E-04	(0.004)	0.00
ϕ_B	0.4	0.80	(0.059)	-0.40	0.43	(0.032)	-0.03
ϕ_C	0.6	0.19	(0.107)	0.41	0.57	(0.111)	0.03
γ	-1.5	-3.30	(0.245)	1.80	-1.55	(0.086)	0.05

smaller when using Chamberlain's instruments, indicating an increase in efficiency. Those results are robust in Columns (9)-(11) where we introduce preference heterogeneity with random coefficients on the valuation of the expected attribute γ.

Table 2 displays the estimation results in the data where only information types B and C are present while we estimate the data with an over-specified model that includes all three types. As the information type A is missing in the DGP, we expect that the estimate of its fraction ϕ_A is not significantly different from zero. Columns (3)-(5) show that the regular instruments fail to solve the endogeneity issue. Similar to Table 1, the estimates of the intercept, information parameters ϕ and valuation of the expected attribute γ are heavily biased. Using Chamberlain's instruments reduces significantly the bias and results in smaller standard errors. Especially in column (6), the estimated fraction ϕ_A is negligible in scale and not statistically significant. This indicates the absence of the type A in the DGP. By plotting the GMM objective function, we observe that the regular instruments lead to a problematic objective function with multiple local minima, as shown in Figure 1. However, the GMM objective function becomes well-behaved when using Chamberlain's instruments, as depicted in Figure 2, further supports the effectiveness of Chamberlain's instruments.

7.3 Simulation of Moment Inequality Approach

For ease of computation, we slightly change the setup in 7.1 and simulate one single dataset, consisting of $T = 5000$ markets and $J = 10$ products, that only contains two information types. The indirect utility now excludes the price variable and is defined as

$$u_{ijt} = \beta_0 + \beta_x x_{jt} + \gamma g_{jt} + \xi_{jt} + \epsilon_{ijt}.$$

Table 3: Simulation Results with $\kappa_O = A$

	β_0	β_x	γ	Num. Obs.
True values	1	1	-1.5	
95% confidence intervals	1	[0.33,1.17]	[-1.73,-1.03]	4 995

The product characteristic is normally distributed $x_{jt} \overset{\text{iid}}{\sim} N(4,1)$. The uncertain attribute is assumed to be product-market specific $g_{ijt} = g_{jt}$ and equals the sum of the three observed information variables $g_{jt} = k_{1jt} + k_{2jt} + k_{3jt}$. The three information variables are independent of each other. The first information variable is normally distributed $k_{1jt} \overset{\text{iid}}{\sim} N(5,1)$. The second follows a Student distribution of degree of freedom 50 i.e., $k_{2jt} \overset{\text{iid}}{\sim} T(50)$. The third follows a normal distribution $k_{3jt} \overset{\text{iid}}{\sim} N(0, 0.5)$ The demand shifters are uniformly distributed $\xi_{jt} \overset{\text{iid}}{\sim} U(-1,1)$. The demand shocks follow an EVT1 distribution. The preference parameters are $\theta = (\beta_0, \beta_x, \gamma)' = (1, 1, -1.5)'$.

There are two information types in the simulations, indexed by $\kappa = A, B$, where the true information set of each type is $\mathcal{W}_{At} = \{k_{1jt}\}, \mathcal{W}_{Bt} = \{k_{1jt}, k_{2jt}, k_{3jt}\}$. The proportion of each information type is $\phi_A = 0.40, \phi_B = 0.60$. Each information type forms the expectations $g^e_{\kappa jt}$ as follows: $g^e_{Ajt} = k_{1jt} + \mathbb{E}[k_{2jt}|k_{1jt}] + \mathbb{E}[k_{3jt}|k_{1jt}]$, and $g^e_{Bjt} = k_{1jt} + k_{2jt} + k_{3jt}$. Here, the information type B corresponds to a perfect foresight type.

Furthermore, we can allow the researcher to have a partially mis-specified information set. Here, we assume that the researcher specifies that the information set of each type is $\tilde{\mathcal{W}}_{At} = \{k_{1jt}\}, \tilde{\mathcal{W}}_{Bt} = \{k_{1jt}, k_{2jt}\}$. Thus, the researcher does not observe all the information that individuals use to form expectations. Specifically, the information type B has more information about the options than researchers assume they have. The mixture model in Section 4 will be biased in this setting. We show that our moment inequality approach can deliver informative and precise bounds for the parameters of interest.

We estimate the moment inequalities in (15) using the information type A as the observed type κ_O defined in Assumption 5. We select observations where type A's prediction errors e_{Ajt} are above the 90th percentile $q_{0.9}$ (and below the 10th percentile $q_{0.1}$) of its empirical distribution. The selection instruments are hence $Z^S_{\max jt} = \{e_{Ajt} \geq q_{0.9}\}$ and $Z^S_{\min jt} = \{e_{Ajt} \leq q_{0.1}\}$.

Table 3 presents the simulation results. Our moment inequality approach provides informative and precise bounds. We discuss the impact of selecting different observations on the estimated confidence regions in Appendix C.

8 Empirical Applications

We bring the finite mixture framework to the setting of two empirical papers: Dickstein and Morales (2018) (with individual-level data) and Grigolon et al. (2018) (with market-level data).

8.1 What Do Exporters Know?

Dickstein and Morales (2018) studies the information structure of exporters in Chilean manufacturing sectors using moment inequalities based on a minimal information set shared by all exporters. Rather than assuming a minimal information set, our key identification Assumption 1 requires that the observed list of information variables allows us to construct every information type among the exporters.

We show that the finite mixture method can provide estimates close to the original paper's confidence sets without using moment inequalities. Second, by estimating the fraction of each information type in the mixture, we can directly infer the distribution of exporters' information types. We can not only answer the question "What do *all* exporters know?" but also estimate the prevalence of different information types in the data. Finally, through this application, we demonstrate our method can be easily adapted to cases where richer individual-level data is available.

8.1.1 Model

A firm i located in Chile makes the decision d_{ijt} on whether to export to a foreign country j in year t. Her decision is defined as

$$d_{ijt} \equiv \mathbb{1}\left\{\mathbb{E}\left[\pi_{ijt}|\mathcal{J}_{ijt}, dist_j, \nu_{ijt}\right] \geq 0\right\},$$

where π_{ijt} is the profit of firm i from exporting to country j in year t, $\mathcal{J}_{ijt}$ is the information set that firm i uses in year t to predict the potential revenue r_{ijt} from exporting to country j, $dist_j$ is the distance between Chile and a destination country j, and ν_{ijt} is the demand shock. The profit is defined as the difference between the revenue $\eta^{-1}r_{ijt}$, scaled by the demand elasticity η calibrated at $\eta = 5$, and the fixed cost of export f_{ijt}. The fixed cost $f_{ijt} = \beta_0 + \beta_1 dist_j + \nu_{ijt}$ is specified as a linear function of the distance $dist_j$ and the demand shock ν_{ijt}. We can rewrite the decision rule above as

$$d_{ijt} \equiv \mathbb{1}\left\{\eta^{-1}\mathbb{E}\left[r_{ijt}|\mathcal{J}_{ijt}\right] - \beta_0 - \beta_1 dist_j - \nu_{ijt} \geq 0\right\}.$$

The demand shock is assumed to follow a normal distribution $\nu_{ijt}|(\mathcal{J}_{ijt}, dist_j) \overset{\text{iid}}{\sim} N(0, \sigma^2)$, which allows us to obtain a probit form of the probability that firm i exports to a destination j in year t, conditional on the information set $\mathcal{J}_{ijt}$ and the distance $dist_j$ as follows

$$\Pr(d_{ijt} = 1|\mathcal{J}_{ijt}, dist_j; \theta) = \Phi\left[\sigma^{-1}\left(\eta^{-1}\mathbb{E}\left[r_{ijt}|\mathcal{J}_{ijt}\right] - \beta_0 - \beta_1 dist_j\right)\right], \tag{18}$$

where $\Phi(\cdot)$ denotes the CDF of the standard normal distribution and the parameter of interest is $\theta = (\beta_0, \beta_1, \sigma)'$.

The export revenue r_{ijt} takes the role of the uncertain attribute g_{ijt} in our model. However, we do not observe the revenue r_{ijt} of a firm i if she does not export to the destination j in year t. To deal with the issue, Dickstein and Morales (2018) further specifies the potential revenue as a function of the observed domestic revenue r_{iht} of firm i in year t:

$$r_{ijt} = \alpha_{jt} r_{iht} + e_{ijt},$$

where the revenue shifter α_{jt} is a sufficient statistic of how destination-specific supply and demand factors rescale the value of the domestic revenue to the (counterfactual) potential export revenue r_{ijt}, and e_{ijt} is the error term. We follow their approach and estimate the revenue shifter using data on firms that have exported to the destination j in year t. The paper further assumes that the unobserved error term has a zero conditional mean $\mathbb{E}_{jt}\left[e_{ijt}|\mathcal{J}_{ijt}, r_{iht}, f_{ijt}\right] = 0$. Consequently, we take $\alpha_{jt} r_{iht}$ as the uncertain attribute g_{ijt} in our model.

Dickstein and Morales (2018) discuss the role of five observed information variables: distance $dist_j$, last year's domestic revenue r_{iht-1}, last year's aggregate revenue R_{jt-1} from all firms that have exported to the destination j, last year's revenue shifter α_{jt-1}, and the number of firms N_{jt-1} that have exported to the destination j in year $t-1$. They assume all exporters know the first three information variables and construct the minimal information set $\mathcal{J}^m_{ijt} = \{dist_j, r_{iht-1}, R_{jt-1}\}$. We retain their minimal information set and use the last two variables α_{jt-1} and N_{jt-1} to construct the information types. We construct $K_3 = 4$ information types, indexed by $\kappa = A, B, C, D$, where the information set of each type is $\mathcal{W}_A = \{dist_j, r_{iht-1}, R_{jt-1}\}$, $\mathcal{W}_B = \{dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1}\}$, $\mathcal{W}_C = \{dist_j, r_{iht-1}, R_{jt-1}, N_{jt-1}\}$, and $\mathcal{W}_D = \{dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1}, N_{jt-1}\}$, respectively.

8.1.2 Estimation

The data used for estimation is an unbalanced panel of $N = 266$ Chilean firms' decisions d_{ijt} in the food and chemical product sectors from 1996 to 2005, i.e., $T = 10$. The destination choice set contains $J^c = 22$ countries in the chemical sector and $J^f = 34$ countries in the food sector. On average, 38% of the firms have at least exported to one country in a given year. Each firm exports to $4-5$ countries on average in the chemical sector and $6-7$ in the food sector. To summarize, in each sector in year t, we observe all firms' export decision d_{ijt}, their domestic revenue r_{iht}, the export revenue r_{ijt} of firms who have exported to the destination j and the list of information variables $\mathcal{K}_{ijt} = (dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1}, N_{jt-1})$.

Given individual-level data, we separately estimate the export decision in each sector with the maximum likelihood. In this context, the likelihood function (16) becomes

$$\mathcal{L}(\Theta|d,\mathcal{K},dist) = \prod_{ijt} \left[\Pr(d_{ijt}=1|\mathcal{K}_{ijt},dist_j;\Theta)\right]^{d_{ijt}} \left[1-\Pr(d_{ijt}=1|\mathcal{K}_{ijt},dist_j;\Theta)\right]^{1-d_{ijt}},$$

and the choice probability $\Pr(d_{ijt}=1|\mathcal{K}_{ijt},dist_j;\Theta)$ is specified as

$$\Pr(d_{ijt}=1|\mathcal{K}_{ijt},dist_j;\Theta) = \sum_{\kappa}^{K_3} \phi_\kappa \Phi\left[\sigma^{-1}\left(\eta^{-1}\mathbb{E}\left[\alpha_{jt}r_{iht}|\mathcal{W}_\kappa\right] - \beta_0 - \beta_1 dist_j\right)\right], \tag{19}$$

where the fraction parameters satisfy the constraint $\sum_{\kappa}^{K_3}\phi_\kappa = 1; \phi_\kappa \in [0,1]$ and the parameter vector is $\Theta = (\beta_0, \beta_1, \sigma, \phi_A, \phi_B, \phi_C, \phi_D)'$.

When bringing the likelihood function (19) to the data, we need to calculate the type-specific expectations of the potential export revenue $\mathbb{E}\left[\alpha_{jt}r_{iht}|\mathcal{W}_\kappa\right]$. However, we do not know the shape of the conditional expectation operator.[17] We calculate the conditional expectation with a non-linear exponential specification following Dickstein and Morales (2018). For instance, the expectation of the potential export revenue in information type A, i.e., the minimal information type, is

$$\mathbb{E}\left[\alpha_{jt}r_{iht}|\mathcal{W}_A\right] = \exp\left(\gamma_{A1}\ln(dist_j) + \gamma_{A2}\log(r_{iht-1}) + \gamma_{A3}\log(R_{jt-1})\right), \tag{20}$$

where we first run a non-linear regression of the ex-post export revenue $\alpha_{jt}r_{iht}$ on information variables $\ln(dist_j), \log(r_{iht-1}), \log(R_{jt-1})$ and obtain the estimated coefficients $\hat{\gamma}_{A1}, \hat{\gamma}_{A2}, \hat{\gamma}_{A3}$. Then we predict the unobserved expectation as

$$\hat{\mathbb{E}}\left[\alpha_{jt}r_{iht}|\mathcal{W}_A\right] = \exp(\hat{\gamma}_{A1}\ln(dist_j) + \hat{\gamma}_{A2}\log(r_{iht-1}) + \hat{\gamma}_{A3}\log(R_{jt-1})).$$

We repeat this calculation for each information type κ. The only difference between our calculation and Dickstein and Morales (2018) is that we do not include any intercept term $\gamma_{\kappa 0}$ in (20).[18] We find an exponential shape by plotting the correlation between the predicted attribute $\alpha_{jt}r_{iht}$ and the information variables $\mathcal{W}_{\kappa t}$, which justifies that the functional form

[17]In the simulation exercise, we assume the two information variables are independent and the type-specific expectations reduce to a single information variable with an additional constant, which allows us to know correctly the conditional expectations.

[18]Indeed, when estimating Equation (20) with intercept γ_{A0}, we find that the estimated coefficients $\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3$ are positive but all smaller than 1, while the estimated intercept $\hat{\gamma}_0$ is negative and much larger in absolute value (around -5). This issue exists in the estimation for all information types, which makes the predicted expectations $\hat{\mathbb{E}}\left[\alpha_{jt}r_{iht}|\mathcal{W}_\kappa\right]$ close to 0 for any information type κ, making them indistinguishable.

Table 4: Parameter Estimates for Entry Decision in the Chemical Sector

Estimator	σ	β_0	β_1	ϕ_A	ϕ_B	ϕ_C	ϕ_D
Panel A. Parameter Estimates							
(1) Perfect foresight	1,038.6	745.3	1,087.8				
	(413)	(291)	(422)				
(2) Minimal information	395.6	298.3	447.1				
	(111)	(73)	(129)				
(3) Finite Mixture	124.7	95.5	210.7	0.15	1.17E-12	0.56	0.29
	(11)	(11)	(25)	(0.048)	(0.092)	(0.078)	(–)
Panel B. Confidence Intervals							
Methods robust to unobserved information structure							
(4) Moment inequality	[85.1,115.9]	[62.8,81.1]	[142.5,194.2]				
(5) Finite Mixture	[104.1,145.4]	[73.6,117.5]	[162.5,259]				
Simple plug-in estimators							
(6) Perfect foresight	[229.3,1848.0]	[175.6,1314.9]	[260.4,1915.1]				
(7) Minimal information	[178.1,613.0]	[154.8,441.7]	[193.6,700.6]				

we use to estimate the unobserved conditional expectations should be able to minimize the approximation error.

8.1.3 Results

Table 4 presents the estimation results of the export model in the chemical sector. Rows (1) and (2) replicate the results from Dickstein and Morales (2018) using simple plug-in estimators. The estimator in row (1) assumes all exporters have perfect foresight while the one in row (2) assumes all exporters use the minimal information set. Our estimation results using the finite mixture model with four information types are shown in row (3). From the estimated fractions ϕ_κ, we observe a significant share of information type A, which uses the minimal information set, confirming the findings in Dickstein and Morales (2018). Furthermore, we observe that firms are more likely to include the number of exporters N_{jt-1} in their information set, compared to only using the minimal information set ($\hat{\phi}_A = 0.15 < \hat{\phi}_C = 0.56$). Additionally, exporters are unlikely to use the revenue shifter α_{jt-1} combined with the minimal information set, as $\hat{\phi}_B = 1.17E - 12$, which is negligible and not statistically significant. Intuitively, information about the number of total exporters is easier to obtain and use than the revenue shifter. When a firm knows the minimal information set and wants to enhance it, their first choice might be to include the number of exporters rather than calculating the revenue shifter. However, when a firm knows both the minimal information and the number of exporters, they are likely to further include the revenue shifter ($\hat{\phi}_D = 0.29$).

Panel B of Table 4 displays the confidence intervals for firms' fixed cost estimates $\hat{\sigma}, \hat{\beta}_0, \hat{\beta}_1$. Row (4) borrows the confidence intervals estimated by Dickstein and Morales (2018) using moment inequalities. We further calculate the confidence intervals for our finite mixture model in row (5) and for the two simple plug-in estimators in rows (6) and (7). We find that the simple plug-in estimators result in large and significantly different confidence intervals compared to the moment inequality estimates in row (4). However, our finite mixture confidence interval in row (5) is much more precise and overlap with those obtained using moment inequalities. Figure 3 illustrates this graphically.

8.2 Consumer Valuation of Fuel Costs

Grigolon et al. (2018) studies the car purchase decision in the EU market. The authors use a demand model that is rich in consumer heterogeneity to obtain consistent estimates of consumers' valuation of fuel costs. Specifically, they allow preference heterogeneity using random coefficients and driving behavior (mileage) heterogeneity using an empirical distribution. They provide quantitative evidence that consumers undervalue their expected fuel cost when purchasing a vehicle. The paper adds to a large literature following Hausman (1979) identifying the responsiveness of energy-consuming durables to energy expenses. The literature has, so far, relied on a plug-in approach. The researcher specifies what consumers expect about future energy consumption at the time of the purchase and plugs in the expectation in the purchase decision model. If consumers form expectations differently than what the researcher specified, these types of models suffer from the bias of misspecified information structures discussed in Section 3.

Therefore, we extend Grigolon et al. (2018) and provide a more detailed analysis of consumers' valuation of the uncertain future fuel costs. Specifically, we investigate the information variables consumers use to form their expectations of fuel costs. We recast the problem in our framework and estimate a finite mixture of observed information types. This allows us to obtain estimates of consumers' valuation of their expected fuel costs consistent with a much richer information structure and estimate which information types are prevalent in the data.

8.2.1 Model

A consumer i decides whether to purchase a car model j with engine variant k in market (year-country) t. For simplicity, we omit the market subscript t. Her decision utility is defined as

$$u_{ijk} = x_{jk}\beta_i^x - \alpha_i(p_{jk} + \gamma G_{ijk}) + \xi_{jk} + \epsilon_{ijk},$$

where each car is defined based on its model j and engine type k, x_{jk} is a vector of observed car characteristics, p_{jk} is the price, G_{ijk} is consumer i's expected fuel cost, ξ_{ij} is an unobserved product attribute and ϵ_{ijk} is the idiosyncratic valuation for the car, modeled as an EVT1 random variable. The vector β_i^x is the consumer-specific coefficients on the car characteristics, α_i is the marginal utility of income and the expected fuel cost G_{ijk} accounts for individuals' mileage heterogeneity. Specifically, it represents consumer i's present discounted value of expected future fuel costs for the car model j with engine variant k as

$$G_{ijk} = \rho\beta_i^m e_{jk} g_k,$$

where $\rho \equiv \sum_{s=1}^{S}(1+r)^{-s}$ is the capitalization coefficient that depends on the lifetime S of the car and the interest rate r, β_i^m is consumer i's annual mileage, e_{jk} is the fuel consumption of the car and g_k is the fuel price. Finally, the parameter γ measures consumers' future valuation. If $\gamma < 1$, consumers undervalue future payoffs G_{ijk} relative to the current payoff p_{jk}.

Grigolon et al. (2018) assumes that consumers have perfect foresight of future fuel costs. We relax this assumption and investigate what information consumers use to form their expectations G_{ijk}. We assume consumers know their mileage β_i^m, and we aim to estimate ρ. We take the observed fuel efficiency e_{jk} and fuel prices g_k as the information variables and construct a mixture model of $K_3 = 3$ information types, indexed by $\kappa = A, B, C$, where the information set of each type is $\mathcal{W}_A = \{e_{jk}\}$, $\mathcal{W}_B = \{g_k\}$, and $\mathcal{W}_C = \{e_{jk}, g_k\}$, respectively. We denote the product $e_{jk}g_k$ as the uncertain attribute g_{jk} in our model. The expectations of each information type is $g_{\kappa jk}^e = \mathbb{E}[e_{jk}g_k|\mathcal{W}_\kappa]$. The information type C corresponds to the perfect foresight type that uses both fuel efficiency and fuel price to predict their fuel costs, as specified in Grigolon et al. (2018).

Under these assumptions, we can compute the predicted market share for model j with engine k as

$$s_{jk}(\xi;\rho,\Theta) = \sum_{\kappa=1}^{K_3}\phi_\kappa\left[\int_\beta \frac{\exp(x_{jk}\beta_i^x - \alpha_i p_{jk} - \alpha_i\gamma\rho\beta_i^m g_{\kappa jk}^e + \xi_{jk})}{1+\sum_{j,k}\exp(x_{jk}\beta_i^x - \alpha_i p_{jk} - \alpha_i\gamma\rho\beta_i^m g_{\kappa jk}^e + \xi_{jk})} dF_\beta(\beta;\theta)\right], \quad (21)$$

where the vector of random coefficients $\beta_i = (\beta_i^x, \alpha_i, \gamma\rho\beta_i^m)'$ is assumed to be independent of the taste shock ϵ_{ijt} and follows a distribution $F_\beta(\beta;\theta)$ where θ are means and (co)variance parameters to be estimated. The parameter vector of interest is $\Theta = (\theta', \phi_A, \phi_B, \phi_C)'$.

8.2.2 Estimation

The data used to estimate the market share system in Equation (21) is a panel of T markets, defined as country-year combinations. The vector of product attributes in x_{jkt} includes horsepower, size, and height of the car, whether the car is produced in a foreign country, and a diesel dummy interacted by country dummy variables.

Following Grigolon et al. (2018), we further restrict the random coefficients with

$$\beta_i^x = \bar{\beta}^x + \Sigma^x \nu_i^x,$$

where $\bar{\beta}^x$ is the vector of mean valuations and Σ^x is assumed to be a diagonal matrix with the vector of standard deviations σ^x on the diagonal and ν_i^x follows a standard normal distribution. The marginal utility of income is specified as inversely proportional to the observed market's income level y_t, i.e., $\alpha_i = \alpha / y_t$. Individual mileage β_i^m is drawn from the observed empirical mileage distribution. Given those restrictions, the parameter of interest reduces to $\Theta = (\bar{\beta}^{x'}, \Sigma^{x'}, \alpha, \gamma\rho, \phi_A, \phi_B, \phi_C)'$.

Finally, the unobserved quality ξ_{jkt} is assumed to be linearly additive as

$$\xi_{jkt} = \xi_j + \xi_t + \tilde{\xi}_{jkt},$$

where ξ_j are model-specific fixed effects and ξ_t are market-specific fixed effects modeled as country-specific fixed effects interacted with a time trend and a squared time trend. The model is then estimated with GMM using the conditional moment restrictions

$$\mathbb{E}\left[\tilde{\xi}_{jkt} | z_t\right] = 0,$$

where z_t is a vector of instruments.

To deal with the unobserved shape of the conditional expectation operator $g^e_{\kappa jk} = \mathbb{E}[e_{jk} g_k | \mathcal{W}_\kappa]$, we use the random forest prediction as an approximation. This specification should be able to minimize the approximation error since it is flexible enough to capture the observed non-linear relationship between information variables and the expected fuel costs, as displayed in Figure 3.

We first estimate the mixture model in Equation (21) without random coefficients, i.e., $\beta_i^x = \bar{\beta}^x$ using the standard two-step GMM. In the first stage, our instruments z_t include observed product attributes x_{jkt}, cost shifters, BLP instruments and the realized fuel costs g_{jkt}. Then we follow Reynaert and Verboven (2014) to calculate the approximated Chamberlain's instruments for the fraction parameters ϕ_κ and the undervaluation coefficient γ based on first stage estimates and instruments. In the second stage, we replace the realized fuel costs g_{jkt} by those Chamberlain's instruments in z_t.

When allowing preference heterogeneity, the regular two-step GMM fails to deliver precise estimates in our mixture model (21). Indeed, our fraction parameters can be considered as additional "random coefficients" that characterize the information heterogeneity. As analyzed in Reynaert and Verboven (2014), it is difficult to precisely estimate many random coefficient parameters. Our simulation exercise in Section 7 shows that the quality of the estimates is sensitive to the strength of the instruments. Besides, the strength of the non-linear Chamberlain's instruments depends on the first-stage estimates. To deal with this interdependence dilemma and improve the efficiency of the estimates, we adopt a one-step estimator that continuously updates the Chamberlain's instruments, following Bourreau et al. (2021). When minimizing the GMM objective function in the outer loop, we employ Newton's method with a numerical gradient. Our estimates are robust to the alternative gradient-free Nelder-Mead simplex method used in Bourreau et al. (2021).

8.2.3 Results

Table 5 presents the estimation results of the demand model. The first three columns replicate results from Table 3 in Grigolon et al. (2018) for comparison with our findings displayed in columns (4)-(6). Our main result is that our estimates reject the perfect foresight assumption used in the literature. Specifically, the estimated fraction of the perfect foresight consumer type ϕ_C is negligible and not statistically significant across the three demand specifications. Moreover, we find that more than half of the consumers use only fuel efficiency to predict their future fuel costs, while the remainder rely solely on fuel price. This finding is intuitive, as fuel efficiency is a direct attribute of the car and easy to observe and use, whereas future fuel prices are uncertain and harder to predict. Finally, our results show that the estimated undervaluation coefficient γ in Grigolon et al. (2018) has a large upward bias. Indeed, our mixture model estimates $\hat{\gamma} = 0.23$ is much lower than the 0.91 estimated under the perfect foresight model. This suggests that accounting for consumers' unobserved information about fuel costs is important in this setting. These findings align with survey findings discussed in Levinson and Sager (2023), pointing out that consumers' ex-post and ex-ante fuel costs differ substantially.

Table 5: Parameter Estimates for Alternative Demand Models

	Table 3 in Grigolon et al. (2018)			Finite Mixture		
	Logit (1)	RC Logit I (2)	RC Logit II (3)	Logit (4)	RC Logit I (5)	RC Logit II (6)
Panel A. Mean Valuations						
Price/inc. (α)	-4.52	-6.22	-5.33	-2.44	-4.45	-4.97
	(0.19)	(0.22)	(0.21)	(0.21)	(0.21)	(0.28)
Fuel costs/inc. ($\alpha\gamma\rho$)	-39.03	-46.48	-47.11	-12.73	-18.54	-11.07
	(1.41)	(0.94)	(9.22)	(0.33)	(0.64)	(0.50)
Power (kW/100)	2.28	2.6	0.25	1.14	2.27	-1.87
	(0.14)	(0.17)	(0.61)	(0.14)	(0.14)	(0.17)
Size (cm^2/10k)	13.25	16.69	16.77	14.36	15.79	14.39
	(0.44)	(0.48)	(2.02)	(0.41)	(0.42)	(1.37)
Height (cm/100)	3.00	4.45	5.19	3.05	3.87	5.59
	(0.30)	(0.32)	(0.33)	(0.26)	(0.28)	(0.18)
Foreign	-0.83	-0.75	-0.89	-0.94	-0.84	-1.09
	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.05)
Panel B. Fractions of Information Types						
Fuel efficiency (ϕ_A)				0.68	0.66	0.51
				(0.02)	(0.02)	(0.04)
Fuel price (ϕ_B)				0.32	0.34	0.39
				(0.03)	(0.03)	(0.04)
Perfect foresight (ϕ_C)				0.00	0.00	0.10
				(–)	(–)	(–)
Panel C. Standard Deviations of Valuations						
Power (kW/100)			1.95			2.70
			(0.25)			(0.18)
Size (cm^2/10k)			4.31			6.04
			(2.04)			(1.21)
Foreign			0.49			1.56
			(0.43)			(0.06)
Mileage distribution	No	Yes	Yes	No	Yes	Yes
Panel D. Valuations of Future Fuel Costs						
Fuel costs/price ($\gamma\rho$)	8.63	7.47	8.84	5.22	4.17	2.23
	(0.55)	(0.24)	(1.77)	(–)	(–)	(–)
Future val. γ ($r = 6\%$)	0.89	0.77	0.91	0.54	0.43	0.23
	(0.06)	(0.02)	(0.18)	(–)	(–)	(–)

References

Abaluck, Jason and Giovanni Compiani (2020) "A Method to Estimate Discrete Choice Models that is Robust to Consumer Search," Working Paper 26849, National Bureau of Economic Research, 10.3386/w26849.

Andrews, Donald W. K. and Xiaoxia Shi (2013) "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81 (2), 609–666, https://doi.org/10.3982/ECTA9370.

Andrews, Isaiah, Jonathan Roth, and Ariel Pakes (2023) "Inference for Linear Conditional Moment Inequalities," *The Review of Economic Studies*, 90 (6), 2763–2791, 10.1093/restud/rdad004.

Arcidiacono, Peter, V. Joseph Hotz, Arnaud Maurel, and Teresa Romano (2020) "Ex Ante Returns and Occupational Choice," *Journal of Political Economy*, 128 (12), 4475–4522, 10.1086/710559.

Aucejo, Esteban and Jonathan James (2021) "The Path to College Education: The Role of Math and Verbal Skills," *Journal of Political Economy*, 129 (10), 2905–2946, 10.1086/715417.

Barseghyan, Levon, Maura Coughlin, Francesca Molinari, and Joshua C. Teitelbaum (2021) "Heterogeneous Choice Sets and Preferences," *Econometrica*, 89 (5), 2015–2048, https://doi.org/10.3982/ECTA17448.

Bergemann, Dirk, Benjamin Brooks, and Stephen Morris (2022) "Counterfactuals with Latent Information," *American Economic Review*, 112 (1), 343–68, 10.1257/aer.20210496.

Berry, Steven and Panle Jia (2010) "Tracing the Woes: An Empirical Analysis of the Airline Industry," *American Economic Journal: Microeconomics*, 2 (3), 1–43, 10.1257/mic.2.3.1.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995) "Automobile Prices in Market Equilibrium," *Econometrica*, 63 (4), 841–890, `http://www.jstor.org/stable/2171802`.

Berry, Steven T. (1994) "Estimating Discrete-Choice Models of Product Differentiation," *The RAND Journal of Economics*, 25 (2), 242–262, `http://www.jstor.org/stable/2555829`.

Berry, Steven T. and Philip A. Haile (2014) "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, 82 (5), 1749–1797, https://doi-org.gorgone.univ-toulouse.fr/10.3982/ECTA9027.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa (2022) "Discretizing Unobserved Heterogeneity," *Econometrica*, 90 (2), 625–643, https://doi.org/10.3982/ECTA15238.

Bourreau, Marc, Yutec Sun, and Frank Verboven (2021) "Market Entry, Fighting Brands, and Tacit Collusion: Evidence from the French Mobile Telecommunications Market," *American Economic Review*, 111 (11), 3459–99, 10.1257/aer.20190540.

Brown, Zach Y. and Jihye Jeon (2024) "Endogenous Information and Simplifying Insurance Choice," *Econometrica*, 92 (3), 881–911, https://doi.org/10.3982/ECTA18555.

Chamberlain, Gary (1987) "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 34 (3), 305–334, https://doi.org/10.1016/0304-4076(87)90015-7.

Ciliberto, Federico and Elie Tamer (2009) "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77 (6), 1791–1828, https://doi-org.gorgone.univ-toulouse.fr/10.3982/ECTA5368.

Compiani, Giovanni and Yuichi Kitamura (2016) "Using mixtures in econometric models: a brief review and some new results," *The Econometrics Journal*, 19 (3), C95–C127, https://doi.org/10.1111/ectj.12068.

Cox, Gregory and Xiaoxia Shi (2023) "Simple Adaptive Size-Exact Testing for Full-Vector and Subvector Inference in Moment Inequality Models," *The Review of Economic Studies*, 90 (1), 201–228, 10.1093/restud/rdac015.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach (2010) "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78 (3), 883–931, https://doi-org.gorgone.univ-toulouse.fr/10.3982/ECTA6551.

Cunha, Flavio, James Heckman, and Salvador Navarro (2005) "Separating uncertainty from heterogeneity in life cycle earnings," *Oxford Economic Papers*, 57 (2), 191–261, 10.1093/oep/gpi019.

Dickstein, Michael J, Jihye Jeon, and Eduardo Morales (2024) "Patient Costs and Physicians' Information," Working Paper 32014, National Bureau of Economic Research, 10.3386/w32014.

Dickstein, Michael J and Eduardo Morales (2018) "What do Exporters Know?," *The Quarterly Journal of Economics*, 133 (4), 1753–1801, 10.1093/qje/qjy015.

D'Haultfœuille, Xavier, Arnaud Maurel, and Yichong Zhang (2018) "Extremal quantile regressions for selection models and the black–white wage gap," *Journal of Econometrics*, 203 (1), 129–142, https://doi.org/10.1016/j.jeconom.2017.11.004.

Einav, Liran, Amy Finkelstein, and Mark R. Cullen (2010) "Estimating Welfare in Insurance Markets Using Variation in Prices," *The Quarterly Journal of Economics*, 125 (3), 877–921, 10.1162/qjec.2010.125.3.877.

Gandhi, Amit, Zhentong Lu, and Xiaoxia Shi (2013) "Estimating demand for differentiated products with error in market shares," CEMMAP Working Paper CWP03/13, Centre for Microdata Methods and Practice (cemmap), London, `https://users.ssc.wisc.edu/~xshi/research/Demand`.

Ghorbanzadeh, Dariush, Philippe Durand, and Luan Jaupi (2017) "A method for the Generate a random sample from a finite mixture distributions," in *CMCGS 2017. 6th Annual International Conference on Computational Mathematics, Computational Geometry & Statistics*, Singapore, Singapore, March, 10.5176/2251-1911_CMCGS17.3.

Grigolon, Laura, Mathias Reynaert, and Frank Verboven (2018) "Consumer Valuation of Fuel Costs and Tax Policy: Evidence from the European Car Market," *American Economic Journal: Economic Policy*, 10 (3), 193–225, 10.1257/pol.20160078.

Hausman, Jerry A. (1979) "Individual Discount Rates and the Purchase and Utilization of Energy-Using Durables," *The Bell Journal of Economics*, 10 (1), 33–54, http://www.jstor.org/stable/3003318.

Houmark, Mikkel Aagaard, Victor Ronda, and Michael Rosholm (2024) "The Nurture of Nature and the Nature of Nurture: How Genes and Investments Interact in the Formation of Skills," *American Economic Review*, 114 (2), 385–425, 10.1257/aer.20220456.

Levinson, Arik and Lutz Sager (2023) "Who Values Future Energy Savings? Evidence from American Drivers," *Journal of the Association of Environmental and Resource Economists*, 10 (3), 717–751.

McLachlan, G.J. and D. Peel (2004) *Finite Mixture Models*, Wiley Series in Probability and Statistics: Wiley, https://books.google.fr/books?id=c2_fAox0DQoC.

Molinari, Francesca (2020) "Microeconometrics with partial identification," in *Handbook of Econometrics, Volume 7A*, 7 of Handbook of Econometrics, 355–486: Elsevier, https://doi.org/10.1016/bs.hoe.2020.05.002.

Porcher, Charly, Eduardo Morales, and Thomas Fujiwara (2024) "Measuring Information Frictions in Migration Decisions: A Revealed-Preference Approach," Working Paper 32413, National Bureau of Economic Research, 10.3386/w32413.

Reynaert, Mathias and Frank Verboven (2014) "Improving the performance of random coefficients demand models: The role of optimal instruments," *Journal of Econometrics*, 179 (1), 83–98, https://doi.org/10.1016/j.jeconom.2013.12.001.

T. Tin Nguyen, Faicel Chamroukhi, Hien D. Nguyen and Geoffrey J. McLachlan (2020) "Approximation by finite mixtures of continuous density functions that vanish at infinity," *Cogent Mathematics & Statistics*, 7 (1), 1750861, 10.1080/25742558.2020.1750861.

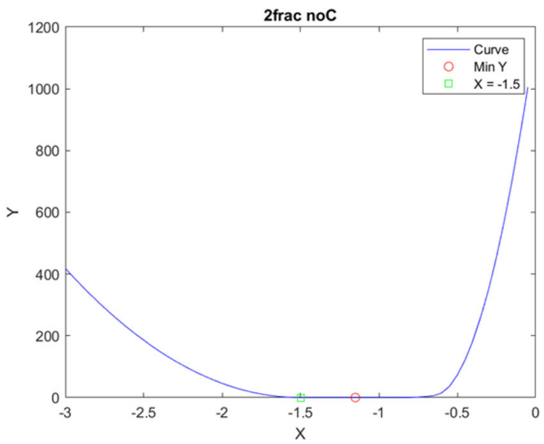
Vatter, Benjamin (2024) "Quality disclosure and regulation: Scoring design in medicare advantage," working paper, https://ssrn.com/abstract=4250361.

Wooldridge, Jefrey M. (2010) *Econometric Analysis of Cross Section and Panel Data*: The MIT Press, http://www.jstor.org/stable/j.ctt5hhcfr.

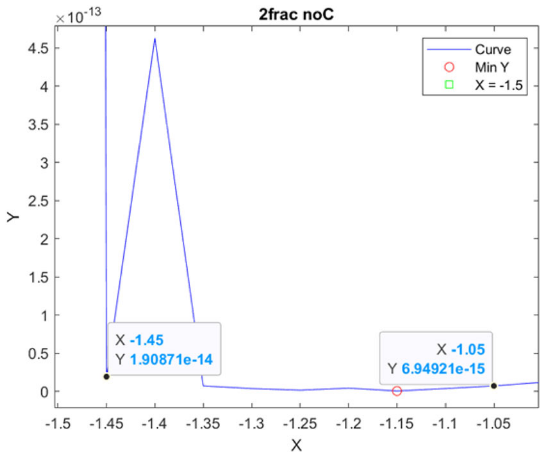
Figures

Figure 1: GMM Objective Function with Regular IVs in the Two-type DGP (B, C included)

(a) Global View



(b) Local View



(c) Global Minimum

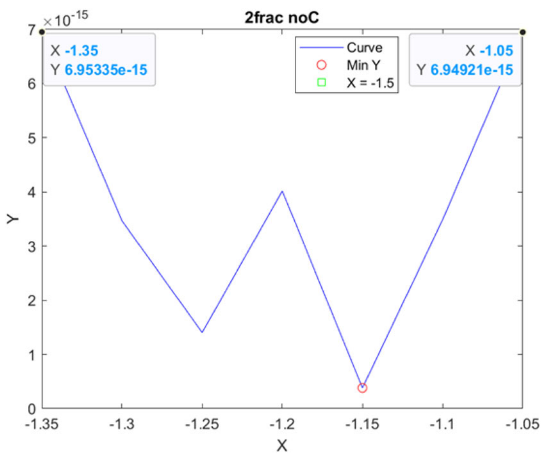
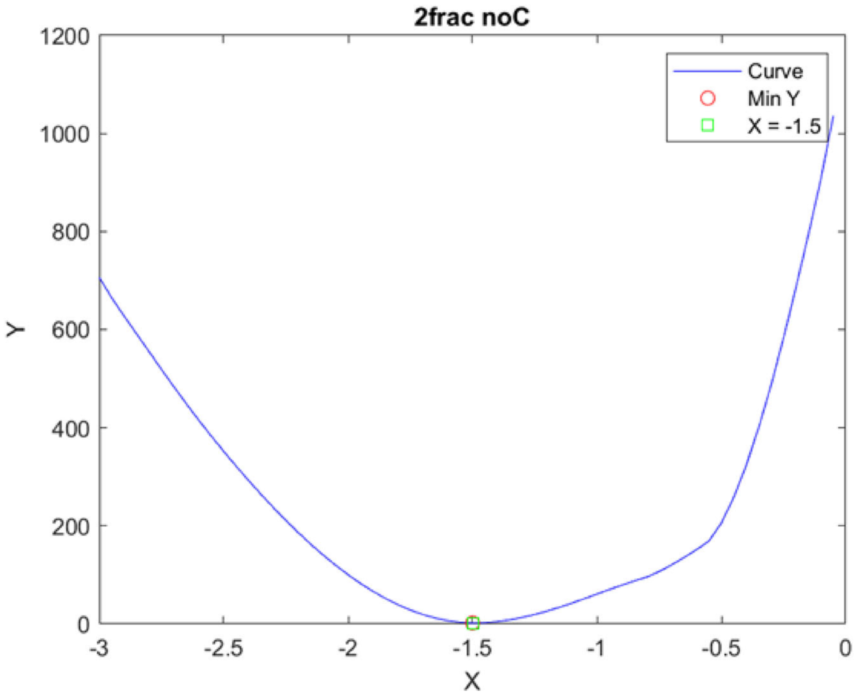


Figure 2: GMM Objective Function with Chamberlain IVs in the Two-type DGP (B, C included)

(a) Global View



(b) Local View

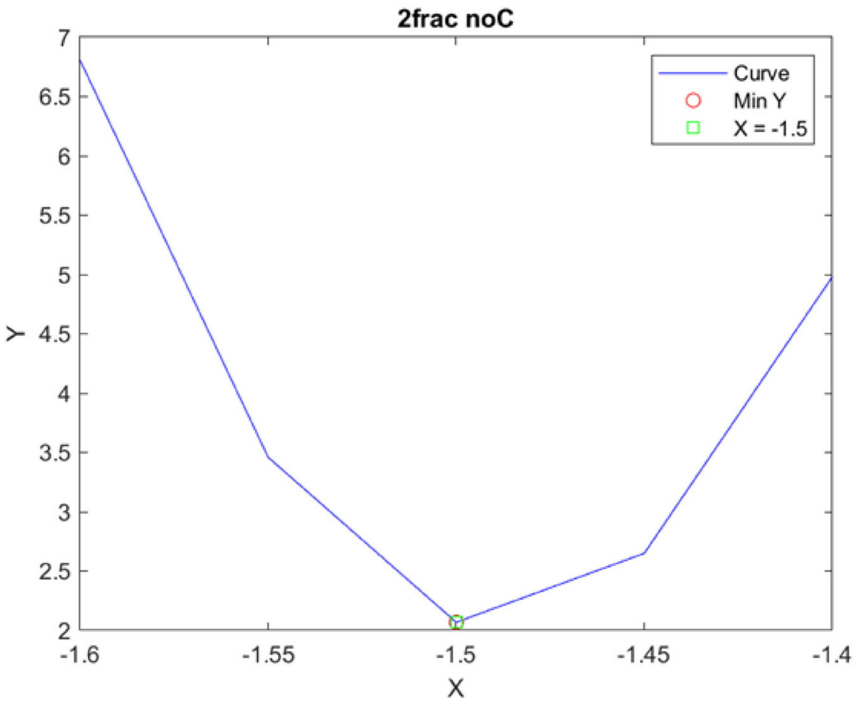


Figure 3: Non-linear Relationship between Information Variables e_{jkt}, g_{kt} and Uncertain Fuel Costs $g_{jkt} = e_{jkt} \times g_{kt}$

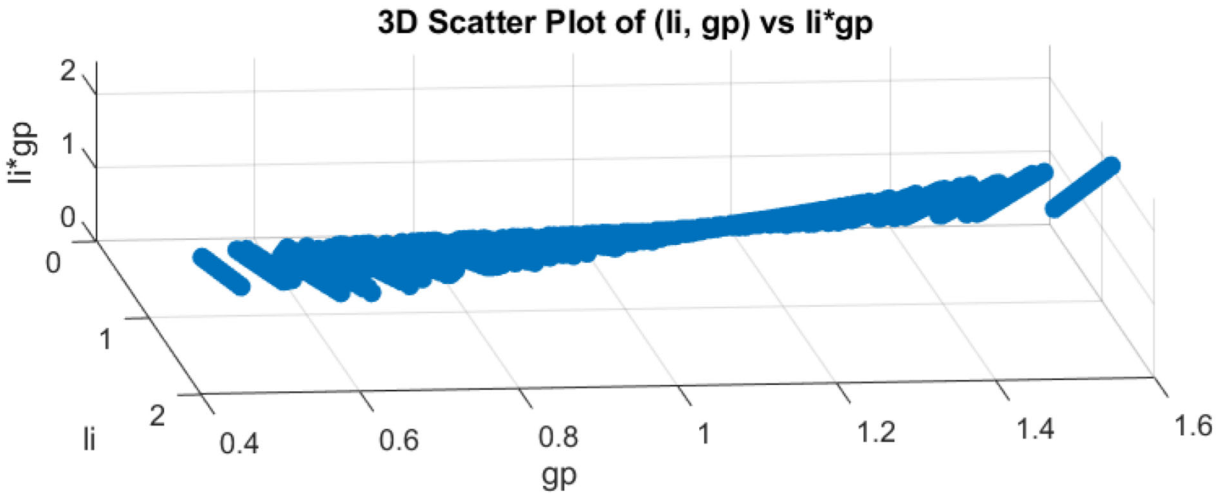


Figure 4: 95% Confidence Intervals for the Chemical Sector

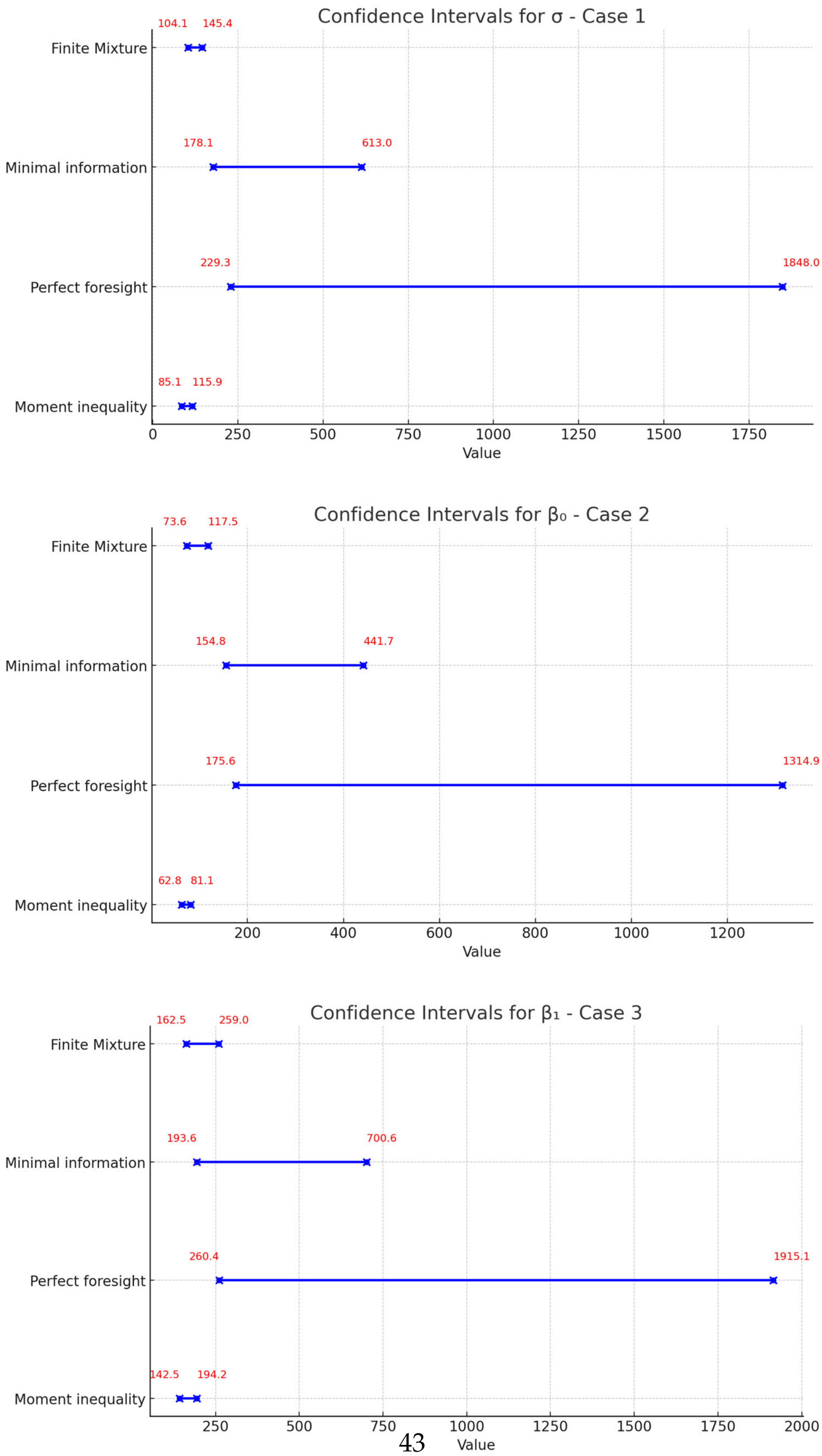
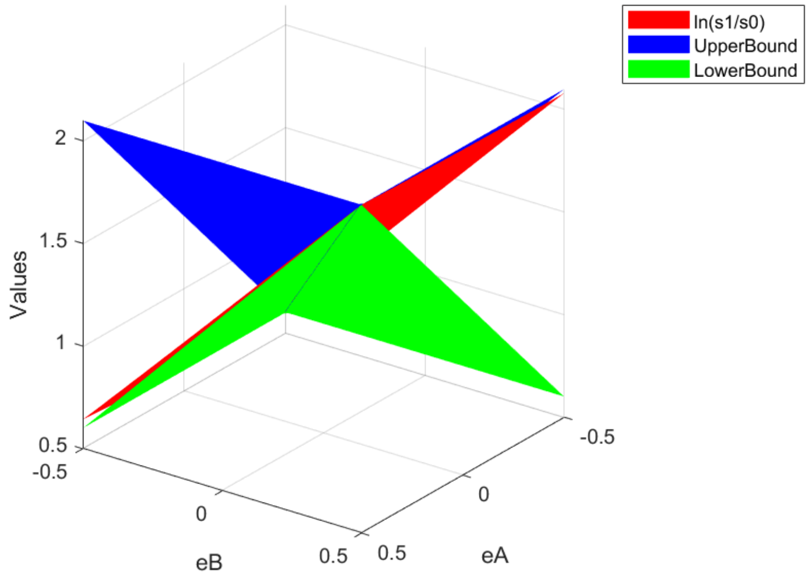
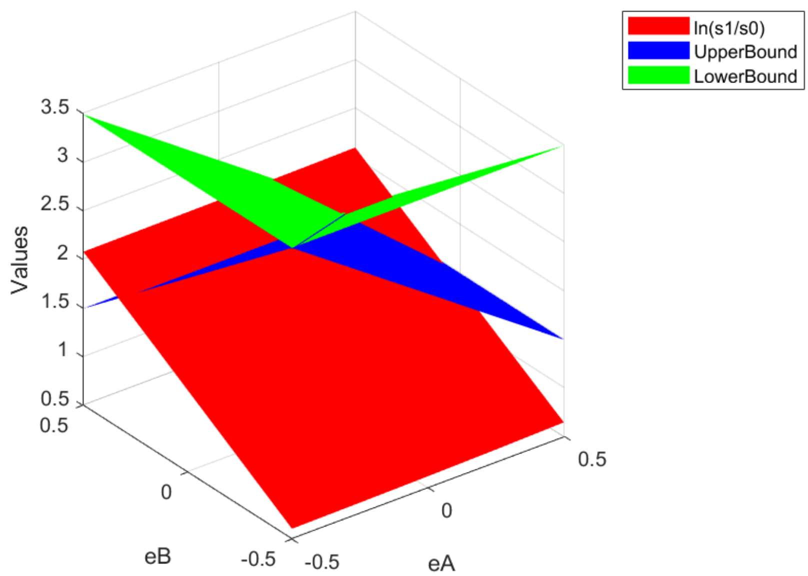


Figure 5: Graphical Illustration of Inequalities

(a) Under True Value of $\gamma = -1.5$



(b) Under False Value of $\gamma = 2$



Appendices

A Proof of Proposition 1

We denote $F^o(\cdot)$ for distributions that can be directly observed or estimated and $f(\cdot)$ for distributions containing any unobserved variables. The data $\{s_t, g_t, \mathcal{K}_t\}_{t\in\mathcal{T}}$ allows to estimate the joint distribution $F^o(s_{jt}, \mathcal{K}_{jt}, g_{ijt})$ across i, j, and t. Without loss of generality, we can write

$$F^o(s_{jt}, \mathcal{K}_{jt}, g_{ijt}) = \int_{g^e_{ijt}} f(d_{ijt}(g^e_{ijt}), \mathcal{K}_{jt}, g_{ijt}) dg^e_{ijt}, \tag{22}$$

where the joint distribution $f(d_{ijt}(g^e_{ijt}), \mathcal{K}_{jt}, g_{ijt})$ involves the unobserved individual choices d_{ijt} that is a function of the unobserved expectations g^e_{ijt}. Equation (22) connects the observed data with the unobserved variable. Using rules of conditional distributions, we have

$$F^o(s_{jt}, \mathcal{K}_{jt}, g_{ijt}) = \int_{g^e_{ijt}} f^y(d_{ijt}|\mathcal{K}_{jt}, g^e_{ijt}, g_{ijt}) \cdot f^y(g_{ijt}|\mathcal{K}_{jt}, g^e_{ijt}) \cdot f^y(g^e_{ijt}|\mathcal{K}_{jt}) \cdot F^o(\mathcal{K}_{jt})\, dg^e_{ijt}, \tag{23}$$

where $F^o(\mathcal{K}_{jt})$ denotes the marginal distribution of $\mathcal{K}_{jt}$ and is observed in the data. Any structure $S^y \equiv \{f^y(d_{ijt}|\mathcal{K}_{jt}, g^e_{ijt}, g_{ijt}), f^y(g_{ijt}|\mathcal{K}_{jt}, g^e_{ijt}), f^y(g^e_{ijt}|\mathcal{K}_{jt})\}$ is admissible provided that it verifies the restrictions imposed in Section 2 and Equation (23). Section 2 implies the logit form of $f^y(d_{ijt}|\mathcal{K}_{jt}, g^e_{ijt}, g_{ijt})$ and we have:

$$f^y(d_{ijt}|\mathcal{K}_{jt}, g^e_{ijt}, g_{ijt}) = \left(\frac{\exp(\gamma g^e_{ijt})}{1+\sum_{j=1}^{J}\exp\left(\gamma g^e_{ijt}\right)}\right)^{d_{ijt}} \left(1 - \frac{\exp(\gamma g^e_{ijt})}{1+\sum_{j=1}^{J}\exp\left(\gamma g^e_{ijt}\right)}\right)^{1-d_{ijt}}. \tag{24}$$

Next, we show that γ is partially identified in a model with stricter assumptions than those in Section 2. The idea is that if we can show partial identification in a more restrictive model, then the model defined in Section 2 is also partially identified. Specifically, our additional assumptions are on the elements of Equation (23):

$$\begin{cases} g_{ijt}^{e} & \sim \mathcal{N}(\mu_{g^e}, \sigma_{g^e}^2), \\ \mathcal{W}_{jt} & = g_{ijt}^{e} + v_{ijt},\ v_{ijt}|g_{ijt}^{e} \sim \mathcal{N}(\mu_v, \sigma_v^2), \\ g_{ijt} & = g_{ijt}^{e} + e_{ijt},\ e_{ijt}|(g_{ijt}^{e}, v_{ijt}) \sim \mathcal{N}(0, \sigma_e^2), \end{cases} \tag{25}$$

where we maintain that expectational errors e_{ijt} have a zero conditional mean, which is a property by construction. Those assumptions allow us to further determine the terms $f^y(g_{ijt}|\mathcal{K}_{jt}, g_{ijt}^e)$ and $f^y(g_{ijt}^e|\mathcal{K}_{jt})$.

First, the distribution of g_{ijt}, given values of $g_{ijt}^e = g$, is fully determined by the distribution of e_{ijt}, independently of $\mathcal{K}_{jt}$. Thus, we have the following normal conditional density:

$$f^y(g_{ijt}|\mathcal{K}_{jt}, g_{ijt}^e) = f^y(g_{ijt}|\mathcal{K}_{jt}, g_{ijt}^e = g) = \frac{1}{\sigma_e^2\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{g_{ijt} - g}{\sigma_e^2}\right)^2\right].$$

Second, by Bayes' rule, we have

$$f^y(g_{ijt}^e|\mathcal{K}_{jt}) = \frac{f(\mathcal{K}_{jt}|g_{ijt}^e)f(g_{ijt}^e)}{f(\mathcal{K}_{jt})},$$

where $f(\mathcal{K}_{jt}|g_{ijt}^e)$ is determined by $\mathcal{K}_{jt}|g_{ijt}^e \sim \mathcal{N}(g_{ijt}^e + \mu_v, \sigma_v^2)$, $f(\mathcal{K}_{jt})$ is obtained with $\int f(\mathcal{K}_{jt}|g_{ijt}^e)dF(g_{ijt}^e)$ where $F(g_{ijt}^e)$ is the CDF of $\mathcal{N}(\mu_{g^e}, \sigma_{g^e}^2)$, and $f(g_{ijt}^e)$ is also determined by that normal distribution. Then we succeed in writing the right-hand side of Equation (23) with the parameter γ and other distribution parameters restricting unobservables $(g_{ijt}^e, v_{ijt}, e_{ijt})$ from the system of assumptions (25).

Recall that our objective is to show that there exist empirical distributions $F^o(s_{jt}, \mathcal{K}_{jt}, g_{ijt})$ for which one can find at least two structures

$$\begin{cases} S^{y_1} \equiv \{\gamma^{y_1}, f^{y_1}(g_{ijt}|\mathcal{K}_{jt}, g_{ijt}^e), f^{y_1}(g_{ijt}^e|\mathcal{K}_{jt})\}, \\ S^{y_2} \equiv \{\gamma^{y_2}, f^{y_2}(g_{ijt}|\mathcal{K}_{jt}, g_{ijt}^e), f^{y_2}(g_{ijt}^e|\mathcal{K}_{jt})\}, \end{cases}$$

that satisfy Equations (23), (24), (25) and $\gamma^{y_1} \neq \gamma^{y_2}$.

Define $g^e_{ijt} = \sigma_{g^e}\tilde{g}^e_{ijt}$ such that $\text{Var}(\tilde{g}^e_{ijt}) = 1$. We can rewrite equation (24) as

$$f^y(d_{ijt}|\mathcal{K}_{jt}, g^e_{ijt}, g_{ijt}) = \left(\frac{\exp(\gamma\sigma_{g^e}\tilde{g}^e_{ijt})}{1+\sum_{j=1}^{J}\exp\left(\gamma\sigma_{g^e}\tilde{g}^e_{ijt}\right)}\right)^{d_{ijt}}\left(1-\frac{\exp(\gamma\sigma_{g^e}\tilde{g}^e_{ijt})}{1+\sum_{j=1}^{J}\exp\left(\gamma\sigma_{g^e}\tilde{g}^e_{ijt}\right)}\right)^{1-d_{ijt}}.$$

We can identify the unique γ if we know σ_{g^e}.

To solve for σ_{g^e}, we can use the system of assumptions (25). We note that g_{ijt} and $\mathcal{K}_{jt}$ are jointly normal as each of them is a sum of normal variables. Their joint distribution can be computed with observed data and reflects the information contained in the system of assumptions (25). We can use the following three moments to summarize the parameters involved in that joint distribution:[19]

$$\begin{cases} \sigma^2_g = \sigma^2_{g^e} + \sigma^2_e, \\ \sigma^2_{\mathcal{K}} = \sigma^2_{g^e} + \sigma^2_v + 2\rho_{vg^e}, \\ \rho_{g\mathcal{K}} = \sigma^2_{g^e} + \rho_{vg^e}. \end{cases}$$

The LHS of this system of equations can be directly observed in the data. This is a linear system with three equations and four unknowns $(\sigma^2_{g^e}, \sigma^2_e, \sigma^2_v, \rho_{vg^e})$, which is under-identified and cannot solve for a unique $\sigma^2_{g^e}$. Consequently, our model in Section 2 cannot be point identified since a more restrictive model with an additional system of assumptions (25) remains partially identified.

B Proof of Theorem 1

We focus on the demand model with market-level data defined in Equation (11). We begin our derivation with an inequality that bounds the ratio of market shares with choice probabilities of two extremal information types.[20]

[19]We use $e|(g^e, v) \sim \mathcal{N}(0, \sigma^2_e)$ to compute $\text{Cov}(v, e) = \mathbb{E}[ve] - \mathbb{E}[v]\,\mathbb{E}[e] = \mathbb{E}[\mathbb{E}[ve|g^e, v]] - \mathbb{E}[v]\,\mathbb{E}[\mathbb{E}[e|g^e, v]] = \mathbb{E}[v\mathbb{E}[e|g^e, v]] - \mathbb{E}[v] \times \mathbb{E}[0] = \mathbb{E}[v \times 0] = 0$. Similarly, $\text{Cov}(g^e, e) = 0$.

[20]The inequality, in general, relates the ratio of sums of sequences to the ratio of individual terms in the sequences.

Lemma 1 *Let ϕ_κ denote the fraction of an information type $\kappa = 1, \ldots, K_3$ in the population, $s_{\kappa jt}$ denote the probability of the information type κ choosing any inside the option j, and $s_{\kappa 0t}$ denote the probability of the information type κ choosing the outside option 0. By definition, $\{\phi_\kappa\}$, $\{s_{\kappa jt}\}$ and $\{s_{\kappa 0t}\}$ are finite sequences of non-negative real numbers. We have:*

$$\min_\kappa \left\{ \frac{s_{\kappa jt}}{s_{\kappa 0t}} \right\} \leq \frac{s_{jt}}{s_{0t}} \leq \max_\kappa \left\{ \frac{s_{\kappa jt}}{s_{\kappa 0t}} \right\}.$$

Proof. Let $M \equiv \max_\kappa \left\{ \frac{\phi_\kappa s_{\kappa jt}}{\phi_\kappa s_{\kappa 0t}} \right\}$ denote the maximum value of the ratio of type-specific choice probabilities. For any information type κ, we have $\phi_\kappa s_{\kappa jt} \leq M \cdot \phi_\kappa s_{\kappa 0t}$. Taking the summation over the type index κ on both sides of that inequality, we have $\sum_{\kappa=1}^{K_3} \phi_\kappa s_{\kappa jt} \leq M \cdot \sum_{\kappa=1}^{K_3} s_{\kappa 0t}$, which implies $\frac{\sum_{\kappa=1}^{K_3} \phi_\kappa s_{\kappa jt}}{\sum_{\kappa=1}^{K_3} \phi_\kappa s_{\kappa 0t}} \leq M \equiv \max_\kappa \left\{ \frac{\phi_\kappa s_{\kappa jt}}{\phi_\kappa s_{\kappa 0t}} \right\}$. By Equation (11), the left-hand side of the latter inequality corresponds to the ratio of market shares for an inside option j and the outside option 0, i.e., $\frac{s_{jt}}{s_{0t}} = \frac{\sum_{\kappa=1}^{K_3} \phi_\kappa s_{\kappa jt}(\delta_t, g_{\kappa t}^e; \theta)}{\sum_{\kappa=1}^{K_3} \phi_\kappa s_{\kappa 0t}(\delta_t, g_{\kappa t}^e; \theta)}$. Thus, we have $\frac{s_{jt}}{s_{0t}} \leq \max_\kappa \left\{ \frac{\phi_\kappa s_{\kappa jt}}{\phi_\kappa s_{\kappa 0t}} \right\} = \max_\kappa \left\{ \frac{s_{\kappa jt}}{s_{\kappa 0t}} \right\}$. The proof is similar for the other side of the inequality. ■

Intuitively, the inequality relaxes the need for a complete specification of the component types and mixing proportions that enter the market shares, instead allowing us to bound parameters with extremal types. These inequalities become equalities when there is no heterogeneity in information between types, suggesting that the bounds tighten in product markets where informational differences play a smaller role, much like the identification at infinity argument in Ciliberto and Tamer (2009).

Next, since the logarithm transformation of ratios is monotone, we can further linearize Lemma 1 as:

$$-\gamma \min_\kappa \{e_{\kappa jt}\} + \gamma g_{jt} + X_{jt}\beta + \xi_{jt} \leq \log \frac{s_{jt}}{s_{0t}} \leq -\gamma \max_\kappa \{e_{\kappa jt}\} + \gamma g_{jt} + X_{jt}\beta + \xi_{jt},$$

which gives the moment inequalities in Equation (13).

C How does Selection Impacts Estimated Confidence Regions?

We estimate the moment inequalities in (15), assuming that both information types A and B are correctly specified by the researchers. To compute the maximum and minimum moment functions, we considered five different scenarios for selecting observations:

1. Only information type A is used in estimation.
2. Only information type B is used in estimation.
3. Both information types A and B are used in estimation.
4. Information type A (B) is used for the maximum (minimum) moment function.
5. Information type B (A) is used for the maximum (minimum) moment function.

In the first scenario, where only type A is used in the estimation, we select observations where type A makes the largest (smallest) errors to compute the maximum (minimum) moment functions. Among the 50 000 observations, 24 968 have type A as the maximum type (with type B as the minimum for the same observations), and 25 032 have type A as the minimum type (with type B as the maximum for the same observations). Since the estimation procedure requires an equal number of observations for the maximum and minimum moment functions, we randomly select 24 968 observations from the 25 032 data points where type A makes the smallest error. The moment inequalities in (15) rewrite as

$$m_{\max}(s_{jt}, g_{jt}, X_{jt}; \theta) = -\log \frac{s_{jt}}{s_{0t}} + \gamma g^{e}_{Ajt} + X_{jt}\beta + \xi_{jt},$$
$$m_{\min}(s_{jt}, g_{jt}, X_{jt}; \theta) = \log \frac{s_{jt}}{s_{0t}} - \gamma g^{e}_{Ajt} - X_{jt}\beta - \xi_{jt}.$$

The second scenario follows the same reasoning but plugs predictions from information type B in the moment functions. The moment inequalities in (15) rewrite as

$$m_{\max}(s_{jt}, g_{jt}, X_{jt}; \theta) = -\log \frac{s_{jt}}{s_{0t}} + \gamma g^e_{Bjt} + X_{jt}\beta + \xi_{jt},$$
$$m_{\min}(s_{jt}, g_{jt}, X_{jt}; \theta) = \log \frac{s_{jt}}{s_{0t}} - \gamma g^e_{Bjt} - X_{jt}\beta - \xi_{jt}.$$

In the third scenario, the entire sample is used for estimation. In our simulation of two types, when type A is the maximum for one observation, type B is necessarily the minimum for that same observation. Thus, each observation can be used twice, once for the maximum and once for the minimum moment functions. The fourth scenario is similar to the third, with each observation used twice. However, the focus is on the 24 968 observations where type A is the maximum type. The moment inequalities in (15) rewrite as

$$m_{\max}(s_{jt}, g_{jt}, X_{jt}; \theta) = -\log \frac{s_{jt}}{s_{0t}} + \gamma g^e_{Ajt} + X_{jt}\beta + \xi_{jt},$$
$$m_{\min}(s_{jt}, g_{jt}, X_{jt}; \theta) = \log \frac{s_{jt}}{s_{0t}} - \gamma g^e_{Bjt} - X_{jt}\beta - \xi_{jt}.$$

The fifth scenario mirrors the fourth but focuses on the 25 032 observations where type B is the maximum type. The moment inequalities in (15) rewrite as

$$m_{\max}(s_{jt}, g_{jt}, X_{jt}; \theta) = -\log \frac{s_{jt}}{s_{0t}} + \gamma g^e_{Bjt} + X_{jt}\beta + \xi_{jt},$$
$$m_{\min}(s_{jt}, g_{jt}, X_{jt}; \theta) = \log \frac{s_{jt}}{s_{0t}} - \gamma g^e_{Ajt} - X_{jt}\beta - \xi_{jt}.$$

After selecting the subsample, we estimate the confidence region for the preference parameter θ using a grid search method following Andrews and Shi (2013). To speed up computation, we fix the intercept $\beta_0 = 1$ and estimate only the parameters β_x and γ. We define a grid of 31 values for each parameter, with β_x ranging from -2 to 3 and γ from -5 to 2. Then we interact with both parameters' grids to form a final grid of $31^2 = 961$ points.

Table 6 displays the simulation results. Comparing scenarios (1) and (2), we find that the subsample using information type A estimates the upper bound for the positive parameter β_x and the lower bound for the negative parameter γ more precisely than the subsample using information type B. When comparing scenarios (1) and (2) with scenario (3), we

Table 6: Simulation Results

	β_x	γ	Num. Obs.
True values	1	-1.5	
95% confidence intervals			
(1) Only A	[0.67, 1.50]	[-1.97, -1.27]	24 968
(2) Only B	[0.67, 2.83]	[-3.60, -1.50]	24 968
(3) Both A, B	[0.83, 1.00]	[-1.73, -1.50]	50 000
(4) Max. A, Min. B	[-1.67, 0.83]	[-1.73, -1.50]	24 968
(5) Max. B, Min. A	[0.83, 1.00]	[-1.73, -1.50]	25 032

observe that using each observation to estimate both moment functions yields the most precise confidence intervals for both parameters β_x, γ.