WORKING PAPERS

N° 1571

March 2025

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 *

March 28, 2025

Abstract

Agents often make choices by forming expectations about attributes, but such expectations are usually unobserved by researchers. We develop two methods for estimating discrete choice models where agents use unobserved heterogeneous information sets to form expectations. Preferences are point-identified using a finite mixture approximation of the unobserved information structure or set-identified with partial information. Both methods apply to individual- and market-level data without imposing strong assumptions on how expectations are formed. We revisit two empirical applications that confirm the importance of accounting for unobserved information: firms' revenue expectations when exporting and consumers' fuel cost expectations when purchasing cars.

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 Christian Bontemps, Arnaud Maurel and seminar participants at Toulouse School of Economics, Vienna University, and APIOC 2024, 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 the attributes of options in their choice set. For example, when choosing health insurance, individuals form expectations about 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). All these settings require the researcher to formulate the expectations agents form about the uncertain attributes of their options to estimate preferences.

Formulating these expectations is difficult in many settings as researchers seldom observe the information set agents use to predict the uncertain attribute of options. Furthermore, we expect agents to differ in the information used when making predictions. In most applications, researchers assume a particular expectation formation process, such as perfect foresight, and use the resulting predicted attributes for estimation. Whenever researchers' formulation of these expectations differs from how agents form expectations, the estimation of preferences is biased (Manski (1993, 2004); Cunha and Heckman (2007);Dickstein and Morales (2018)). Our paper develops two new methods to deal with unobserved and heterogeneous information when estimating preferences. We present simulation results and revisit two empirical applications showing the importance of accounting for unobserved information heterogeneity.

The first method we develop generalizes previous plug-in approaches in which researchers specify a specific form of expectation formation. We propose formulating agent expectations as a finite mixture model of unobserved heterogeneous information types. This method results in point-identified preferences, is consistent when expectations are heterogeneous, is easily implementable with existing choice model estimation methods, and allows us to estimate the prevalence of different information types in the data, thereby substantially expanding the approach of Cunha et al. (2005) (CHN05 hereafter). The key requirement of the finite mixture model is that it requires the researchers to observe an exhaustive set of information variables.

Therefore, we develop a second method that remains useful when researchers observe only the information set of a single type. Because we make weaker assumptions about the information structure, we obtain set-identified preferences. We construct moment inequalities based on the observation that the known type is an extremal information type whose choices bound other types at some realizations of the data, an approach similar in spirit to D'Haultfœuille et al. (2018). Our approach relates to Dickstein and Morales

(2018) (DM18 hereafter). While DM18 rely on the observed minimal information common to all agents, we rely on observing a single information type, thus developing moment inequalities that can be estimated with both individual- and market-level choice data.

The finite mixture method consists of two steps. First, we generate all possible combinations of information variables that agents can use to create a set of discrete information types. For each type in the set, we transform the ex-post product attributes observed by the researcher into a conditional expectation function, maintaining the standard rational expectations assumption in the literature (Ahn and Manski (1993)). Second, we plug the conditional expectations into the choice model and estimate the proportion of agents with each conditional expectation as a finite mixture model. This approach builds on the intuition of Bonhomme et al. (2022) in approximating unobserved continuous heterogeneity with discrete groups.

We prove point-identification of preferences whenever the finite number of types in our model covers all the information types in the data-generating process (DGP). The main assumption is that we can observe every potential information variable agents might use. The key identifying variation comes from (1) how different information variables generate various conditional expectations and (2) how realized choices vary with different realizations of the information variables and associated conditional expectations. While regular plug-in approaches in the literature result in biased preference estimates, the finite mixture plug-in we propose results in a consistent estimator because it allows flexibility in approximating the unobserved information structures.

Our approach has several advantages. The method allows for heterogeneity across information types and directly estimates the fraction of each information type. This contributes relative to Cunha and Heckman (2007), whose approach allows for a trial and error procedure to test information sets against each other. The finite mixture model reveals the distribution of information types in the data, which is often a direct policy interest. Evaluating policy interventions that target information, as in Allcott and Knittel (2019) or Barahona et al. (2023), requires disentangling preference and information heterogeneity. Our method provides a tractable solution to this problem that researchers can employ while using standard choice estimation data and methods. The finite mixture model is compatible with individual- and aggregate-level choice data. Researchers can estimate informational heterogeneity while controlling for attribute endogeneity and random coefficients in preferences, including the preference for the uncertain attribute.

In our second approach, we develop novel moment inequalities that rely on less stringent assumptions than the finite mixture model. In some settings, it can be a strong assumption that researchers can specify all the potential information variables that agents may use. We relax this assumption and prove the set-identification of preferences by bounding the observed market shares by the type-specific choice probabilities of a single information type. We assume that we know the information set of that single type in the data. This allows us to estimate the conditional expectation of the uncertain attribute of this observed type and its prediction errors. Next, we show we can specify an instrument that selects observations in the data where the observed information type makes extreme errors. A feasible instrument is a cut-off for extreme values of the estimated prediction errors of this observed information type. The selection of extremal data points is similar to 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 DM18. We show the identifying power of this method theoretically and in simulations.

We apply the methods developed in this paper to two empirical settings. First, we revisit DM18, which estimates a model of exporters' destination choices. Using the observed ex-post profits, we estimate the conditional expectations that firms form ex-ante about these profits based on combinations of all the information variables in the original dataset. We assume that we capture every potential information type in the DGP and estimate a finite mixture model of four information types. Reassuringly, our information mixture model yields parameter estimates with overlapping confidence intervals relative to those derived from the set identification method in DM18 based on minimal information. We also confirm that every information type receiving a positive weight in the estimation incorporates the minimal information set specified in DM18, with the most prevalent type additionally employing the number of existing exporters in forming expectations about future export profits. As such, our finite mixture model validates the main assumption in DM18 and extends the estimated parameters to include proportions of information types that account for the observed export choices. Moreover, we demonstrate the feasibility of our moment inequality approach in this setting, assuming that at least one exporter type uses the minimal information set. This is different from the main assumption in DM18, which requires the minimal information set to be used by all exporters. We obtain similar confidence intervals from both approaches.

In a second application, we revisit the estimation of consumer fuel cost valuations when making automobile purchases (Grigolon et al. (2018)) (GRV18 hereafter). This literature typically makes strong assumptions about the exact expectations consumers form regarding fuel costs by plugging in a perfect information prediction of fuel costs in the choice model. We apply our model by estimating a conditional expectation function for every possible combination of the underlying information variables that can predict fuel costs. We then specify the model as a finite mixture of all the possible information types and estimate their weights. We find that the estimated valuation for fuel costs in GRV18 is substantially biased upwards, and our results show that we can reject the presence of consumers with full information about fuel costs. We find that a majority of car buyers form fuel cost expectations either based on the fuel efficiency of the cars or based on the fuel type of the car, rarely using both sources of information to accurately predict fuel costs.

Our paper contributes to an empirical literature that models agent beliefs when making choices. DM18 proposes a minimal information set approach with set identification that is expanded in Dickstein et al. (2024) for physicians' prescription choices, and Porcher et al. (2024) for migration. Our paper contributes by making different assumptions on the information structure, allowing us to achieve point identification in a finite mixture approach and to also provide 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) show that 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 using any mental model to form expectations as long as it can be specified as a conditional rational expectation.

Our finite mixture method can be interpreted as a generalization of previous methods that incorporate rational expectations into choice models, see CHN05 and Cunha and Heckman (2007). These methods have found applications in many empirical settings (Houmark et al. (2024);Aucejo and James (2021)). This approach consists of formulating how information enters the decision of agents and separating information from error terms (noise). Across choices, these noise terms should be independent of the information used in decision-making if the researcher correctly specifies the information set. This then allows us to test agents' use of different information sets. Although our approach is conceptually similar, we differ in that we estimate a distribution over the information sets that best fit the data. In a DGP with a single homogeneous information set, our procedure would put all weight on this correct information set because placing weight on alternative information sets would not maximize the data likelihood. Our finite mixture approach is more general in three ways. First, our procedure selects the information sets that best explain the choices. This removes the need for testing specific information sets against each other. Second, our approach allows for heterogeneity in the information sets used by agents through the mixing function. Third, our approach is readily applicable to multinomial choice settings, market-level datasets, and different types of agents' prediction problems as it does not require the explicit specification of the bias stemming from misspecified information.

Our moment inequality approach contributes to the literature by relying on minimal assumptions on the unobserved information structure. Relative to DM18, we develop an approach compatible with individual- and market-level choice data. Bounding aggregate choice probabilities by individual choice probabilities in combination with extremal selection is a method that can be broadly useful in other settings with unobserved individual heterogeneity.

Our paper is structured as follows. Section 2 presents the choice model. Section 3 explains the estimation issues stemming from the unobserved information structure. Section 4 introduces the finite mixture approach. Section 5 introduces the moment inequality approach. Section 6 covers estimation. Section 7 presents the 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 individuals indexed by $i \in \mathcal{N}_t = \{1, \ldots, N_t\}$. Each individual i chooses one option $j \in \mathcal{J} = \{0, 1, \ldots, J\}$ 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 (row) vector of choice characteristics, $g_{ijt} \in \mathbb{R}$ is an uncertain choice attribute whose 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. Extreme value type I (EVT1) distributed in all options j for each individual i in market t. The indirect utility of the outside option is normalized as $u_{i0t} = \epsilon_{i0t}$.

The vector of preference parameters of interest is $\theta = (\beta', \gamma)'$.

Denote agent i's decision variable d_{ijt}, with $d_{ijt} = 1$ if she chooses option j in market t and $d_{ijt} = 0$ otherwise. Agent 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 her 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 have 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 an 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 in the decision-making stage, and they form an expectation g^e_{ijt} of this unknown attribute. We specify the agent'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 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., $\mathcal{W}_{it} = (\{k_{1jt}\}_{j \in \mathcal{J}}, \{k_{2jt}\}_{j \in \mathcal{J}})$) to predict the uncertain attribute, while another individual i' may have $\mathcal{M}_{i'} = \{3\}$ and uses k_{3jt} instead (i.e., $\mathcal{W}_{i't} = (\{k_{3jt}\}_{j \in \mathcal{J}})$). Note that the realized values of the information variables k_{mjt} can vary across options j and markets t.

In the decision-making stage, individual i predicts the uncertain attribute g_{ijt} based on her information set. We model the prediction as follows:

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

$\{\xi_{jt}\}_{j\in\mathcal{J}}, \{\epsilon_{ijt}\}_{j\in\mathcal{J}}$ do not help the agent forecast the uncertain attribute. A corollary of this is that agents' predictions g^e_{ijt} are not correlated with the unobserved attribute ξ_{jt}.[1] 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) and (4), agent i's expected utility from choice j in market t is

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

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 $\delta_{jt} \equiv X_{jt}\beta + \xi_{jt}$, an individual-specific utility term $\mu_{ijt} \equiv \gamma g^e_{ijt}$, and an individual error term ϵ_{ijt}.[2]

Researchers 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 not observing 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, researchers observe the realized values of the uncertain attribute $g_t = \{g_{ijt}\}_{i\in\mathcal{N}_t,j\in\mathcal{J}}$. Besides, researchers typically do not observe the exact content of the information sets $\mathcal{W}_{it}$. Instead, we 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 expectations.

We consider settings where researchers observe individual choices $d_t = \{d_{ijt}\}_{i\in\mathcal{N}_t,j\in\mathcal{J}}$ or aggregate choice data $s_{jt} = \{s_{jt}\}_{j\in\mathcal{J}}$. 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 recover the contents of individuals' information sets. We close this section with two illustrative examples.

Example I (Exporting Firms): Researchers observe the individual-level data $\{d_t, X_t, g_t, \mathcal{K}_t\}_{t\in\mathcal{T}}$. This is the case in DM18 where researchers observe the export decision d_{ijt} of each Chilean firm i 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 the observed attributes X_{jt}, such as the distance between Chile and the destination country. The profit also depends on the export revenue 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

[1]This assumption parallels the standard demand estimation assumption that unobserved preference heterogeneity is not correlated to unobserved product quality. However, in informational settings, it can be argued that the information used to predict attributes might be related to unobserved quality. We find that this potential correlation of two unobservable is difficult to tackle without more data and discuss this further in Section 4.

[2]Our approach is compatible with any other individual unobserved utility terms, such as random coefficients, as long as the unobserved heterogeneity is uncorrelated with the information variables.

information sets $\mathcal{I}_{it}$.[3] 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 (Automobile Purchase): Researchers observe the market-level data $\{s_t, X_t, g_t, \mathcal{K}_t\}_{t\in\mathcal{T}}$, as in Berry (1994) and Berry et al. (1995). For example, in one of our applications, GRV18 uses market-level data to estimate consumers' valuation of fuel costs when purchasing automobiles. 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. This utility depends on the observed attributes X_{jt}, such as the price p_{jt} of the car. It also depends on the fuel costs g_{ijt} that are uncertain to consumers at the decision-making stage. Each consumer i predicts 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. The researcher does not observe car buyers' expectations of fuel costs, g^e_{ijt}, but observes the realized fuel costs g_{ijt} based on mileage, fuel prices, and fuel economy.

3 Estimation Challenges from Unobserved Information Structures

In this section, we discuss the estimation issues that result from misspecifying the information agents use when making choices. We illustrate this with the market-level data setting.[4]

[3] In DM18, 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., $\mathcal{K}_t \subseteq \mathcal{I}_{ijt}$.

[4] See DM18 for a similar discussion for individual-level data.

Bias in Preference Parameters: We focus on γ, the parameter quantifying the valuation for the expected attribute g^e_{ijt}. A primary challenge is that $g^e_{ijt} \equiv \mathbb{E}\left[g_{ijt}|\mathcal{I}_{it}\right]$ remains unobserved for the researcher in the decision utility and depends on each individual's information set $\mathcal{I}_{it}$. A common simplifying assumption is perfect foresight, where g^e_{ijt} is taken to be equal to the actual realization g_{ijt}.

To illustrate the bias, consider a simple demand model with homogeneous agents, i.e., $g^e_{ijt} = g^e_{jt}, g_{ijt} = g_{jt}$, and each agent predicts g_{jt} up to an error $g_{jt} = g^e_{jt} + e_{jt}$. The market shares s_{jt} in Equation (7) are:

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

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

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

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

When assuming perfect foresight, researchers use the last line of Equation (9) to estimate the parameters. However, the perfect foresight assumption introduces a bias through the expectational errors e_{jt}. Indeed, $g_{jt} = g^e_{jt} + e_{jt}$ implies that $\mathbb{E}[e_{jt}|g_{jt}] \neq 0$, and so g_{jt} is correlated to $\tilde{\xi}_{jt}$, which biases the estimation of the parameter γ. Because $\mathrm{Cov}(g_{jt}, e_{jt}) = \mathrm{Var}(e_{jt}) > 0$ there will be an upward bias (in absolute value) in the estimates of γ, which implies an overestimated valuation of expected attributes.

While this resembles a regular measurement error problem as described in Wooldridge (2010), IV methods require potential instruments based on information shifters that encounter two issues. First, any information shifter used as an instrument must be uncorrelated with unobserved information shifters; otherwise, it will be correlated with the expectational error. This means that it is generally difficult to find excluded instruments for information. Agarwal and Somaini (2022) proposes excluded instruments in a different but related setting with unobserved choice sets. However, the existence of these instruments is setting-specific and may not be generally applicable. Second, linear IV approaches are not directly applicable in a setting with individual heterogeneity in information sets. Both our applications deal with individual-specific information sets.

Point and Partial Identification: We discuss what restrictions are needed to achieve point and partial identification with market-level data $\{s_t, X_t, g_t, \mathcal{K}_t\}_{t\in\mathcal{T}}$ 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 DM18 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 must restrict how the distribution of unobserved expectations varies across individuals to point-identify γ, and the rational expectations assumption gives us a valuable set of restrictions.

4 Finite Mixture Model

In this section, 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 use to form their expectations about the attribute. In the second step, we fit the predicted conditional expectations as informational types in a finite mixture model. The conditional expectation functions vary when agents use different information variables to form their expectations, 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 $\mathcal{K}_t = \{(k_{1jt}, \ldots, k_{K_2jt})\}_{j\in\mathcal{J}}$ represents a set of information variables that individuals can use to form their expectations g^e_{ijt}. No other* *information is available to predict g^e_{ijt}.*

This assumption restricts the composition of individuals' information set to observable information variables. This contrasts with the minimal information set assumption in DM18. 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 could use to form their expectations, thereby allowing for a flexible information structure. Our approach also differs from CHN05. Their approach assumes that all individuals share the same set of information variables. They also require that the information factors are mutually independent. In contrast, we assume the researcher can specify all possible information variables and explicitly accommodate heterogeneous information sets across individuals. Our approach also allows for arbitrary correlation among the information variables.

Based on Assumption 1, we can collect all 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 $\mathcal{K}_t$. Each information type, indexed by $\kappa \in \{1, \ldots, K_3\}$, corresponds to a specific information set $\mathcal{W}_{\kappa t}$. 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 an information type using the ex-post value g_{ijt} as the information variable. This discretization allows us to obtain 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:

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

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

Assumption 3 is in line with the previous approaches of DM18 and CHN05.[5] Assumption 3 draws on well-known results in statistics that finite mixtures can approximate any arbitrary distribution under sufficient regularity conditions (McLachlan and Peel, 2004; Ghorbanzadeh et al., 2017; T. Tin Nguyen and McLachlan, 2020). We focus on approximating the unobserved distribution of heterogeneous expectations g^e_{ijt} about choice attributes, while previous approaches such as Berry and Jia (2010), Nevo et al. (2016), and Bonhomme et al. (2022) approximate the unobserved distribution of preference heterogeneity.

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

[5] In applications, 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 $\mathcal{W}_{\kappa t}$, thus minimizing the approximation error of the conditional expectation.

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 plug-in 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.

An important additional benefit is that the framework allows for testing the content of agents' information sets. We can test whether the type κ using the information set $\mathcal{W}_{\kappa t}$ exists if the associated fraction parameter ϕ_κ significantly differs from zero. A positive and significant ϕ_κ implies that the information variables in the set $\mathcal{W}_{\kappa t}$ are used by some individuals to form expectations. This test can be conducted for each information type κ. The estimated information type shares κ could be of direct interest when researchers want to estimate the impact of information intervention policy.

Example (Market-level Data): 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 focus on the case where researchers have aggregate market-level data to show identification. The individual-level data case can be treated similarly.

Proposition 1 *For a given value of θ, 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.*

Because each value of θ is associated with a unique vector of ϕ, identification of θ relies on the invertibility of the market share system as in Berry and Haile (2014).

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

$$\underset{(J+1)\times K_3}{A_t} = \{s_{\kappa jt}\}_{j\in\mathcal{J};\ \kappa=1,\ldots,K_3} = \begin{bmatrix} s_{10t} & s_{20t} & \cdots & s_{K_30t} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1Jt} & s_{2Jt} & \cdots & s_{K_3Jt} \end{bmatrix}, \quad \underset{(J+1)\times 1}{s_t} = \begin{bmatrix} s_{0t} \\ \vdots \\ s_{Jt} \end{bmatrix}.$$

If the square matrix $A_t'A_t$ is invertible, there exists a unique vector of fractions ϕ that solves $A_t\phi = s_t$ in each market t. The sufficient condition to identify information parameters ϕ is that the matrix A_t has full column rank and more rows than columns, i.e., $J+1 \geq K_3$. We need at least $J = K_3 - 1$ options (excluding the outside option) to identify the fraction parameters ϕ. We 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 to achieve identification with $K_3 - 2$ options and an outside good. With knowledge of ϕ, identification of the parameters θ follows Berry and Haile (2014) when the market share system is invertible. ■

The full column rank requirement implies that the variation across columns (i.e., type-specific choice probabilities $s_{\kappa jt}$) is crucial for identifying ϕ. The type-specific expectations $g^e_{\kappa jt}$, generated by the different information variables, are the source of variation across columns. We highlight that the identification argument for ϕ holds within each market t. Thus, variation across markets is unnecessary 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.[6]

[6] In our empirical applications, we assume constant fraction parameters ϕ across markets; hence,

Because the ϕ vector is identified from the information variables, we can extend the model in Equation (11) to include preference heterogeneity, e.g., by introducing random coefficients β_i and γ_i. As in Berry and Haile (2014), identifying the distribution of random coefficients relies on variation across markets in the market shares and choice attributes. As we discussed in Section 2, we require that the information variables are uncorrelated with unobserved preferences heterogeneity.

It is often difficult to distinguish preference and information heterogeneity because one cannot attribute the choice for an option to a strong preference for that option or to specific information about the option. We provide one way to overcome the difficulty of separating preferences and information in our approach. First, we recover the specific effect of information on choice through the rational expectation assumption and the variation in observed information shifters. Second, we disentangle information from preferences because preferences are stable across options and markets, while information variables realize differently across options and markets. See Barseghyan et al. (2021), Molinari (2020), and Agarwal and Somaini (2022) for other examples of disentangling preferences and information.[7]

The semi-parametric finite mixture model developed so far relies on Assumption 1: we need to specify the complete set of information variables that any agent might use and this comes with 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 be unable 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.

5 Moment Inequalities with Extremal Information

We now turn to a partial identification approach to make inference about preferences without specifying a comprehensive set of information variables. We contribute by deriving novel moment inequalities based on evaluating choices when information variables take extreme values. Our approach relaxes Assumption 1 as we do not need the exhaustive set

observations from different markets improve the precision of the estimation by increasing the sample size. 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.

[7]These papers are different from ours because they focus on settings with unobserved choice sets.

of information variables. Relaxing this assumption comes at a cost: the decisions of many unobserved information types cannot be exactly distinguished in the data, and therefore, we only achieve partial identification.

We argue that partial knowledge about the distribution of expectations, in particular its properties at extremums, is sufficient to set-identify the preference parameters θ. This partial identification approach 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. This approach is compatible with market-level data, while previous moment inequality approaches, such as those found in DM18, are compatible with individual-level data.

We start by defining the moment inequalities conditional on the information variables $Z_{jt} = \{\mathcal{W}_{\kappa jt}\}_{\kappa\in\mathcal{K}_t}$:

$$\begin{aligned} \mathbf{m}_{\max}(Z_{jt};\theta) &\geq 0 \\ \mathbf{m}_{\min}(Z_{jt};\theta) &\geq 0, \end{aligned} \tag{12}$$

with

$$\begin{aligned} \mathbf{m}_{\max}(Z_{jt};\theta) &\equiv \mathbb{E}\left[-\log\frac{s_{jt}}{s_{0t}} + \max_{\kappa}\left\{-\gamma e_{\kappa jt}\right\} + \gamma g_{jt} + X_{jt}\beta + \xi_{jt}\middle| Z_{jt}\right], \\ \mathbf{m}_{\min}(Z_{jt};\theta) &\equiv \mathbb{E}\left[\log\frac{s_{jt}}{s_{0t}} - \min_{\kappa}\left\{-\gamma e_{\kappa jt}\right\} - \gamma g_{jt} - X_{jt}\beta - \xi_{jt}\middle| Z_{jt}\right]. \end{aligned} \tag{13}$$

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 parameters of model* (11). *Then $\theta^* \in \Theta_0$.*

Theorem 1 states that the extremal moment inequalities are consistent with the true parameter value θ^*. We provide a proof in Appendix B. To convey the intuition of the proof, we rewrite inequalities (13) as two bounds for the observed log-ratio of market shares:

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

Inequalities (14) show that individuals who commit extremal errors when predicting the uncertain attribute g_{jt} bound the log market share ratio.[8] The two sides of the inequality

[8]The inequalities mirror those found in Gandhi et al. (2023), who employed linear conditional inequalities

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 therefore identifies an upper bound on γ. Conversely, the left-hand side is a lower bound on γ. Appendix Figures A1a and A1b graphically illustrate the power of these inequalities.

Inequalities (14) cannot be estimated directly, since they contain the unobserved terms $\xi_{jt}, \max_{\kappa}\{e_{\kappa jt}\}$ and $\min_{\kappa}\{e_{\kappa jt}\}$. We aim to express some of these terms as observables. To begin, it is helpful to define an extremal type, an individual that has the largest or smallest prediction error:

Definition 1 (Extremal Types) *The extremal 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. $e_{\kappa_{max} jt} \geq e_{\kappa jt} \forall \kappa \neq \kappa_{max}$, i.e., $e_{\kappa_{max} jt} = \max_{\kappa}\{e_{\kappa jt}\}$. Equivalently, the extremal type $\kappa_{\min}$ predicts the uncertain attribute g_{jt} with the smallest error $e_{\kappa jt}$, i.e. $e_{\kappa_{min} jt} \leq e_{\kappa jt} \forall \kappa \neq \kappa_{min}$, i.e. $e_{\kappa_{min} jt} = \min_{\kappa}\{e_{\kappa jt}\}$.*[9]

It is infeasible to recover the full distribution of the extreme errors $(\max_{\kappa}\{e_{\kappa jt}\}, \min_{\kappa}\{e_{\kappa jt}\})$ because that would require knowing the extremal types for each observation g_{jt}. Instead, we assume that there exists an information type $\tilde{\kappa}$ for which the researcher observes the information set $\mathcal{W}_{\tilde{\kappa} jt}$. Based on this information type, the researcher can select instances in the data where the observed type $\tilde{\kappa}$ is the extremal type using a selection condition Z^{S}_{jt}. Additionally, our approach can account for cases where the observed attribute vector X_{jt} contains endogenous attributes, such as prices correlated with the unobserved attribute ξ_{jt}. In these cases, one needs additional instruments for price Z^{P}_{jt}, such as cost shifters, and include them in the instrument set $Z_{jt} = (Z^{P}_{jt}, X_{jt}, \mathcal{W}_{\tilde{\kappa} jt})$. We formalize the assumption as follows:

Assumption 4 ***- Extremal Selection:*** *To set-identify preference with extremal information types, we assume:*

(i) We observe the information set $\mathcal{W}_{\tilde{\kappa}}$ for at least one information type $\tilde{\kappa}$ that exists in the data.

(ii) $Z^{S}_{\max jt}$ is an indicator function that selects data points jt where the observed information type $\tilde{\kappa}$ is the extremal type $\kappa_{\max}$ so that $\mathbb{E}\left[e_{\tilde{\kappa} jt} \cdot Z^{S}_{\max jt} \middle| Z_{jt}\right] = \mathbb{E}\left[e_{\kappa_{\max} jt} \cdot Z^{S}_{\max jt} \middle| Z_{jt}\right]$.

(iii) $Z^{S}_{\min jt}$ is an indicator function that selects data points jt where the observed information type $\tilde{\kappa}$ is the extremal type $\kappa_{\min}$ so that $\mathbb{E}\left[e_{\tilde{\kappa} jt} \cdot Z^{S}_{\min jt} \middle| Z_{jt}\right] = \mathbb{E}\left[e_{\kappa_{\min} jt} \cdot Z^{S}_{\min jt} \middle| Z_{jt}\right]$.

(iv) $\mathbb{E}\left[\xi_{jt} \cdot Z^{S}_{\max jt} \middle| Z_{jt}\right] = 0$ and $\mathbb{E}\left[\xi_{jt} \cdot Z^{S}_{\min jt} \middle| Z_{jt}\right] = 0$.

with market-level data to perform robust inference in settings featuring zero market shares.

[9]This does not mean that $\kappa_{\min}$ has the most accurate expectations, expectation are correct (perfect foresight) when $e_{ijt} = 0$

As an illustration, consider a selection condition $Z^S_{\max jt}$ for maximum errors. One such selection rule is an indicator for the prediction error of the observed type $\tilde{\kappa}$ exceeding a certain threshold a: $Z^S_{\max jt} = \mathbb{1}\{e_{\tilde{\kappa}jt} \geq a_{\max}\}$. The selection condition ensures that the observed type $\tilde{\kappa}$ equals the extremal type $\kappa_{\max}$ for the selected observations. Because $\kappa_{\max}$ makes the largest prediction error, $e_{\kappa_{\max}jt} \geq e_{\kappa jt}, \forall \kappa$, we can recover the unobserved expectations $\mathbb{E}\left[\max_{\kappa}\{e_{\kappa jt}\} \cdot Z^S_{\max jt}\middle| Z_{jt}\right] = \mathbb{E}\left[e_{\kappa_{\max}jt} \cdot Z^S_{\max jt}\middle| Z_{jt}\right] = \mathbb{E}\left[e_{\tilde{\kappa}jt} \cdot Z^S_{\max jt}\middle| Z_{jt}\right]$. A similar threshold instrument can be defined for minimum errors.[10]

In practice, the selection threshold can be constructed from extremal realizations of the prediction errors of type $\tilde{\kappa}$, which we are able to construct. For instance, $a_{\min}$ and $a_{\max}$ can be the 5th of 95th percentile of these prediction errors. Selecting extremal values of the prediction errors of type $\tilde{\kappa}$ makes it likely that we select market shares where the extreme information type is $\tilde{\kappa}$. This selection does introduce a bias-variance trade-off (D'Haultfœuille et al., 2018). A loose selection threshold decreases the variance of the moments, but introduces bias. We discuss this bias-variance trade-off in our simulations in Section 7.

Together, applying these instruments to the moment functions (13) imply the following moment inequalities:

$$\begin{aligned} \mathbf{m}^{\tilde{\kappa}}_{\max}(Z_{jt};\theta) &\equiv \mathbb{E}\left[\left(-\log\frac{s_{jt}}{s_{0t}} + \gamma g^e_{\tilde{\kappa}jt} + X_{jt}\beta + \xi_{jt}\right)\cdot Z^S_{\max jt}\middle| Z_{jt}\right] \geq 0, \\ \mathbf{m}^{\tilde{\kappa}}_{\min}(Z_{jt};\theta) &\equiv \mathbb{E}\left[\left(\log\frac{s_{jt}}{s_{0t}} - \gamma g^e_{\tilde{\kappa}jt} - X_{jt}\beta - \xi_{jt}\right)\cdot Z^S_{\min jt}\middle| Z_{jt}\right] \geq 0, \end{aligned} \tag{15}$$

where we plugin the observed type $\tilde{\kappa}$'s prediction $g^e_{\tilde{\kappa}jt} = g_{jt} - e_{\tilde{\kappa}jt}$.

Example (Automobile Purchase): Consider fuel cost expectations as in Example II and GRV18. Suppose we observe fuel prices and fuel economy, and assume that when fuel prices are extreme, individuals who predict fuel costs solely based on fuel economy commit larger errors than those using other information sets. For the inequalities and selection conditions to work, it suffices to assume that some individuals in the DGP base their predictions only on fuel economy and that, for them, their prediction error becomes larger than any errors made by other individuals when fuel prices exceed a certain threshold.

[10]Alternatively, the selection instrument can be defined with a threshold on the information variable k_{jt}, e.g., $Z^S_{\max jt} = \mathbb{1}\{k_{jt} \geq a_{\max}\}$ since the prediction (and the error) can be expressed as a function of information variables.

6 Estimation

6.1 Finite Mixture Approach

Estimation with Individual-level Data: Given individual-level data, we estimate our parameters of interest with maximum likelihood. The likelihood function is

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

where the choice probability $\Pr(d_{ijt} = 1|\mathcal{I}_{it};\Theta)$ is defined in Equation (10). The estimation proceeds in three steps. First, the researcher states the set of possible information variables and constructs all information types from this set. Second, for each type, the researcher predicts the uncertain attribute.[11] Third, the researcher constructs the likelihood using the predicted uncertain attributes for each information type.

Estimation with Market-level Data: We extend the GMM approach based on Berry et al. (1995), which imposes mean independence of the structural errors ξ_{jt} relative to a set of instruments: $\mathbb{E}\left[\xi_{jt}|Z_{jt}\right] = 0$, 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}:

$$\begin{aligned} &\min_{\Theta} \xi(\Theta)'Z'WZ\xi(\Theta) \\ &\text{subject to } \sum_{\kappa=1}^{K_3} \phi_\kappa = 1, \phi_\kappa \in [0,1], \end{aligned}$$

where W is the optimal weighting matrix, the vectors and matrices are stacked over all markets t.

We assume that the observed product characteristics X_{jt} and information variables k_{mjt}

[11] In applications, we do not know the expectation form $\mathbb{E}[\cdot]$. In practice, we plot the joint and conditional distribution between the ex-post attribute and information variables to learn about the potential shape of the conditional expectations $g^e_{\kappa jt}$. We then use non-parametric methods or polynomial approximations to the functional form.

are mean independent of the structural error.[12] Our identification condition writes

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

While the information variables are obvious candidates as instruments, they perform poorly in simulations and applications because they do not capture the nonlinearity through which information impacts shares through the predicted attribute. At the same time, not all information variables are necessarily relevant as the researcher might include variables not used by any individual.

We find that in applications, Chamberlain's approximation of optimal instruments (Chamberlain, 1987) works 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 the instruments Z_{jt}, including 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 the derivative of the structural error with respect to that irrelevant parameter ϕ_κ to be asymptotically zero. In other words, Chamberlain's instruments delete irrelevant types based on the data by assigning negligible weights to moments associated with non-existing types while favoring other moments related to information that affects market shares. This helps smooth the objective function and ensures consistent estimates.

Reynaert and Verboven (2014) highlight the performance of Chamberlain's instruments in identifying non-linear parameters and provide a procedure to compute an empirical approximation of these instruments. We follow their 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.

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

6.2 Moment Inequalities Approach

The estimation of the moment inequality approach proceeds in three steps. First, we specify the information set $\mathcal{W}_{\tilde{\kappa}t}$ for the information type $\tilde{\kappa}$ and estimate her predictions $g^e_{\tilde{\kappa}jt} = \mathbb{E}[g_{jt}|\mathcal{W}_{\tilde{\kappa}t}]$ and the associated prediction errors $e_{\tilde{\kappa}jt}$. Second, we form the instrument Z^S_{jt} that selects observations where the prediction errors $e_{\tilde{\kappa}jt}$ reach their extreme values. This instrument choice involves a tradeoff. A loose selection criterion that retains much data risks possible misspecification whenever $\tilde{\kappa}$ is not the extremal type. 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).[13]

7 Simulations

To illustrate the power of our approaches, we conduct simulations of both the finite mixture and moment-inequalities approach. We focus on settings with market-level data.

7.1 Setup

Our simulation is based on the example described in Section 4.1 with three information types. We perform a Monte Carlo exercise with 1,000 datasets, each consisting of $T = 25$ markets and $J = 10$ products. The indirect utility is defined as

$$u^e_{ijt} = \beta_0 + \beta_x x_{jt} - \alpha p_{jt} + \gamma g^e_{jt} + \xi_{jt} + \epsilon_{ijt}.$$

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 with a cost shifter $z_{jt} \overset{\text{iid}}{\sim} U(0, 1)$ and a 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

[13] Andrews et al. (2023) and Cox and Shi (2023) provide alternative inference procedures in settings with linear conditional moment inequalities, but applying Andrews and Shi (2013) has been sufficiently powerful to obtain reasonable results in our simulations.

Table 1: Information Types in the Simulation

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

Notes. This table summarizes the values of expectations $g^e_{\kappa jt} \equiv \mathbb{E}[g_{jt}|\mathcal{W}_{\kappa jt}]$, expectational errors $e_{\kappa jt} \equiv g_{jt} - g^e_{\kappa jt}$ and the variance of expectations errors $\mathrm{Var}(e_{\kappa jt})$ of each information type $\kappa = A, B, C$ specified in our simulation. The information sets are $\mathcal{W}_{At} = \{k_{1jt}\}$, $\mathcal{W}_{Bt} = \{k_{2jt}\}$, and $\mathcal{W}_{Ct} = \{k_{1jt}, k_{2jt}\}$. The true value is specified as $g_{jt} = k_{1jt} + k_{2jt}$.

expectation $\mathbb{E}[k_{2jt}] = \exp(0.5)$ and variance $\mathrm{Var}[k_{2jt}] = e(e-1)$ 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 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 $\mathrm{Var}(e_{\kappa jt})$ are in Table 1.

We also provide simulations where we introduce preference heterogeneity through 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 $\gamma_i \overset{\text{iid}}{\sim} N(-1.5, 0.5)$. 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 19.4%.

Finally, we simulate a third simulation with datasets that only contain 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 when there are fewer information types in the data than specified. The average market share of the outside option is 17.6% in such datasets.

Table 2: 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

Notes. All estimates are from 1,000 Monte Carlo simulations and the DGP includes three information types $\kappa = A, B, C$. Columns (3) to (8) are from a DGP where there is no individual heterogeneity in the distribution of preference parameters while columns (9) to (11) are from a DGP where individuals' valuation of the expected attribute follows a normal distribution with mean γ and standard deviation σ_γ. Regular instruments are the intercept, product characteristics x_{jt}, prices p_{jt} and two information variables k_{1jt}, k_{2jt}. Chamberlain's IV replaces the latter two information variables in the instrument set.

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 as regular instruments constant, x_{jt}, p_{jt}, k_{1jt} and , k_{2jt} in the first stage. We calculate the approximated Chamberlain's instruments for parameters γ and ϕ based on the first stage estimates following Reynaert and Verboven (2014). Then, we replace the information variables k_{1jt}, k_{2jt} in the set of regular instruments with Chamberlain's instruments in the second stage. [14]

Table 2 displays the estimation results of a Monte Carlo simulation over the 1,000 datasets when 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. In Columns (9)-(11), we precisely estimate the information type shares and the random coefficients on the valuation of the expected attribute γ.

Table 3 displays the estimation results when the DGP includes two information types B

[14] When estimating the dataset with preference heterogeneity, we find that the first-stage estimates are usually heavily biased due to the lack of strong instruments, affecting the quality of Chamberlain's instruments in the second stage. To address this issue, we implement a three-step procedure by constructing Chamberlain's instrument twice.

Table 3: 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

Notes. All estimates are from 1,000 Monte Carlo simulations and the DGP includes two information types $\kappa = B, C$. Regular instruments are the intercept, product characteristics x_{jt}, prices p_{jt} and two information variables k_{1jt}, k_{2jt}. Chamberlain's IV replaces the latter two information variables in the instrument set.

and C, but we include A, B and C as potential information types to test if we can successfully estimate ϕ_A to equal zero. Columns (3)-(5) show that the regular instruments, similar to Table 2, are substantially biased. Using Chamberlain's instruments shrinks the bias and results in smaller standard errors. The estimated fraction ϕ_A is negligible in scale and has a very small standard error, so we can statistically exclude the importance of Type A in the DGP.[15]

7.3 Simulation of Moment Inequality Approach

For ease of computation, we modify the setup in 7.1 and simulate a single dataset consisting of $T = 2,500$ markets and $J = 10$ products. The indirect utility now excludes the price variable and is defined as

$$u^e_{ijt} = \beta_0 + \beta_x x_{jt} + \gamma g^e_{jt} + \xi_{jt} + \epsilon_{ijt}.$$

The data-generating process follows the standard three-type finite mixture setting, except that we swap the fractions of types A and B, setting $\phi_A = 0.5$ and $\phi_B = 0.15$. We estimate the moment inequalities in (15) using the information type A as the observed type $\tilde{\kappa}$ defined in Assumption 4. We select observations where type A's prediction errors

[15]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 Appendix Figure A2. However, the GMM objective function becomes well-behaved when using Chamberlain's instruments, as depicted in Appendix Figure A2, further supports the effectiveness of Chamberlain's instruments.

Table 4: Simulation Results with Different Selection Thresholds for $\tilde{\kappa} = A$

	β_x (1)	γ (2)	# Obs. (3)	# Misclass (4)	# Pts in CI (5)
True value	1	-1.5			
a					
0.9	[-0.14,1.94]	[-2.20,-0.15]	5 000	0	285
0.8	[0.14,1.77]	[-2.09,-0.48]	10 000	0	199
0.7	[0.14,1.66]	[-1.98,-0.48]	15 000	478	164
0.6	[0.25,1.61]	[-1.88,-0.58]	20 000	2 869	126
0.5	[0.36,1.49]	[-1.77,-0.69]	25 000	5 369	89

Notes. All estimates are from a large dataset of 25,000 observations. The data-generating process includes three information types $\kappa = A, B, C$ and we specify A as the observed extreme type $\tilde{\kappa}$ in the estimation. The instruments are the intercept, x_{jt}, and estimated expectations of the extreme type $\hat{g}^e_{Ajt}$. We present results for different selection thresholds a from 0.9 to 0.5. The selected samples' sizes are reported in column (3). Among the selected observations, we count the number of cases where their $\hat{g}^e_{Ajt}$ are not the most extreme of all type-specific predictions $\hat{g}^e_{\kappa jt}, \kappa = A, B, C$. This is reported in column (4). Both numbers of observations (3) and misclassification (4) are the sum of data points constructing the max and min inequalities (e.g., at 90% percentiles, we have 2,500 obs. for the max side and another 2,500 obs for the min side). In our simulation, the number of misclassified data points used in the estimation all comes from the max side of the inequality. For ease of computation, we do not estimate the intercept and fix it at its true value. We define a grid of 40 values for the other two parameters, with β_x ranging from -0.2 to 2 with a step of 0.055 and γ from -2.2 to -0.1 with a step of 0.0575, whose interactions form a final grid of $40^2 + 1 = 1601$ points. Extreme points of the 95% confidence set are reported in square brackets. The number of points $(\hat{\beta}_x, \hat{\gamma})$ inside the 95% confidence set is reported in column (5).

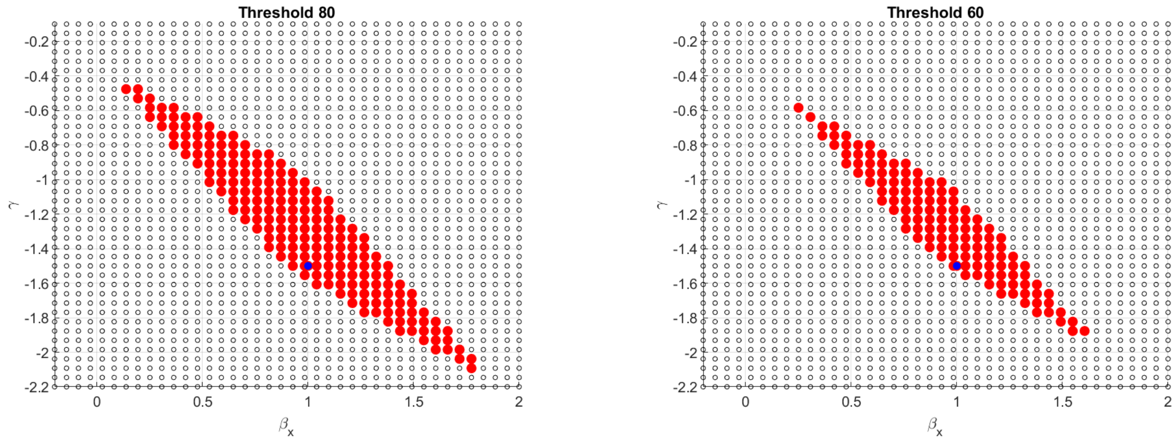
e_{Ajt} are above the a-th quantile q_a (and conversely below the $(1-a)$-th quantile q_{1-a}) of its empirical distribution, with $a \in \{0.9, 0.8, \ldots, 0.5\}$. The selection conditions are defined as $Z^S_{\max jt} = \mathbb{1}\{e_{Ajt} \geq q_a\}$ and $Z^S_{\min jt} = \mathbb{1}\{e_{Ajt} \leq q_{1-a}\}$.

The moment inequalities in (12) rely on knowledge of the prediction error $e_{\tilde{\kappa}jt}$ for the extreme information type $\tilde{\kappa}$. In practice, the true extreme type may vary across observations and the selection instrument might misclassify observations where other types than $\tilde{\kappa}$. are extremal. Such misclassification introduces noise into the moment conditions, which weakens their ability to correctly detect violations and may lead to conservative confidence regions. Intuitively, the looser the selection threshold a, the greater the risk that the observed type is not sufficiently extreme relative to other types. On the other hand, including more observations increases the effective sample size, which improves statistical precision.

Our simulation setting allows us to observe the predictions of all information types when generating the data. This allows us to assess the misclassification errors and the bias-variance trade-offs of our threshold instrument.

Table 4 presents the simulation results. Our moment inequality approach yields informative and precise confidence bounds. We experiment with different selection thresholds to illustrate the bias-variance trade-off inherent in our framework. Column (4) of Table 4 reports the number of misclassified observations included in the estimation

Figure 1: 95% Confidence Regions across Different Selection Thresholds ($\tilde{\kappa} = A$)



Notes. This figure plots the 95% confidence regions (in red) for two representative selection thresholds $a = 0.8$ and $a = 0.6$. The blue point indicates the true value. The confidence region shrinks when the sample size increases from $a = 0.8$ to $a = 0.6$. Although the sample with $a = 0.6$ contains around 14% misclassified observations, the estimated confidence region remains valid.

sample. As expected, we find that as the sample size increases, the number of misclassified observations also increases, but their share, in general, remains small relative to the total estimation sample.

We find that, as the selection threshold decreases (i.e., more observations are included), the size of the confidence region continues to shrink, indicating that the sample size effect dominates the misclassification effect, and the inference is not overly conservative. Until a threshold of $a = 0.8$ there is no misclassification in the simulated data. When the threshold is reduced to $a = 0.7$, although there are around 3% of misclassified observations, the confidence region remains nearly identical to the one obtained with $a = 0.8$. When we decrease the threshold to $a = 0.5$, the full sample is used, and the misclassified observations account for over 20% of the estimation sample. Only in this case does the misclassification problem dominate and result in the exclusion of the true parameter from the confidence region.

Figure 1 illustrates the bias-variance trade-off by graphically showing the evolution of the confidence region. Supplemental evidence is also provided in Figure A3 that plots the estimated bounds for each parameter β_x and γ across different thresholds. This exercise demonstrates that our moment inequality approach can deliver informative and valid inference, provided that (1) the sample size grows sufficiently fast relative to the number of misclassified observations, and (2) the share of misclassified observations remains bounded within the estimation sample.

8 Empirical Applications

We revisit two empirical papers where unobserved information structures play an important role. First, we revisit DM18, which uses individual-level data, and apply both our finite mixture and moment inequality approach. Second, we revisit the market-level data setting of GRV18 and apply our finite mixture approach.

8.1 What Do Exporters Know?

DM18 studies the information structure of exporters in Chilean manufacturing sectors using moment inequalities based on a minimal information set shared by all exporters. On the contrary, our key identification assumption in (1) requires that the observed list of information variables allows us to construct every information type among the exporters. We can directly infer the distribution of exporters' information types by estimating the fraction of each information type in the mixture. We can not only answer the question "What do *all* exporters know?" but also estimate the fraction of exporters with different information sets. In our moment inequality approach, we assume that an exporter type exists that uses the minimal information set, but we do not have to assume the minimal information set is shared among all exporters.

8.1.1 Model

Firm i decides on whether to export, d_{ijt}, to a foreign country j in year t:

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

where π_{ijt} is the profit of firm i from exporting to country j in year t, $\mathcal{I}_{ijt}$ 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 from the destination country j, and ν_{ijt} is a demand shock. The profit equals 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

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

The demand shock follows a normal distribution $\nu_{ijt}|(\mathcal{J}_{ijt}, dist_j) \overset{\text{iid}}{\sim} N(0, \sigma^2)$. The probability that firm i exports to a destination j in year t writes as:

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

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, DM18 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. Under the assumption that $\mathbb{E}_{jt}\left[e_{ijt}|\mathcal{I}_{ijt}, r_{iht}, f_{ijt}\right] = 0.$, $\alpha_{jt} r_{iht}$ equals the uncertain attribute g_{ijt} in our model.

DM18 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 $\mathcal{I}^m_{ijt} = \{dist_j, r_{iht-1}, R_{jt-1}\}$. We retain their minimal information set and use α_{jt-1} and N_{jt-1} to construct additional 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 with Finite Mixture Approach

DM18 employ an unbalanced panel of $N = 266$ unique Chilean firms' decisions to export in the food and chemical product sectors from 1996 to 2005, i.e., $T = 10$. We assume that firms make binary decisions to export to any of each of $J^c = 22$ countries in the chemical sector and $J^f = 34$ countries in the food sector. In each sector and year t, we observe the export decisions of all firms d_{ijt}, their domestic revenue r_{iht}, the export revenue r_{ijt} of firms who have exported to the destination j, and a list of information variables $\mathcal{K}_{ijt} = (dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1}, N_{jt-1})$.

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 constraints $\sum_{\kappa}^{K_3}\phi_\kappa = 1; \phi_\kappa \in [0,1]$ 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 $\mathbb{E}\left[\alpha_{jt}r_{iht}|\mathcal{W}_\kappa\right]$. We calculate the conditional expectation with a non-linear exponential specification following DM18. For instance, the expectation of the potential export revenue in information type A, i.e., the minimal information type, is

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

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

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

We repeat this calculation for each information type κ. The only difference between our calculation and DM18 is that we do not include any intercept term $\gamma_{\kappa 0}$ in (20).[16]

[16]Indeed, when estimating Equation (20) with intercept γ_{A0}, we find that the predicted expectations $\hat{\mathbb{E}}\left[\alpha_{jt}r_{iht}|\mathcal{W}_\kappa\right]$ are close to 0 for any information type κ, making them indistinguishable. Without the intercept,

8.1.3 Estimation with Moment Inequalities

Following the binary choice model in DM18, we estimate the standard deviation $\sigma > 0$ of the demand shock ν_{ijt}, and we abstract from any unobserved product attribute ξ_{jt}. We assume that ν_{ijt} follows an EVT1 distribution instead of a normal distribution. This allows us to directly adapt our inequality specification in Equation (15) as follows:

$$\mathbf{m}^{\tilde{\kappa}}_{\max}(Z_{ijt};\theta) \equiv \mathbb{E}\left[\left(-\log\frac{s_{ijt}}{1-s_{ijt}} + \sigma^{-1}\left(\eta^{-1}\mathbb{E}[r_{ijt} \mid \mathcal{I}^{\tilde{\kappa}}_{ijt}] - \beta_0 - \beta_1 dist_j\right)\right)\cdot Z^S_{\max\, ijt}\middle| Z_{ijt}\right],$$
$$\mathbf{m}^{\tilde{\kappa}}_{\min}(Z_{ijt};\theta) \equiv \mathbb{E}\left[\left(\log\frac{s_{ijt}}{1-s_{ijt}} - \sigma^{-1}\left(\eta^{-1}\mathbb{E}[r_{ijt} \mid \mathcal{I}^{\tilde{\kappa}}_{ijt}] - \beta_0 - \beta_1 dist_j\right)\right)\cdot Z^S_{\min\, ijt}\middle| Z_{ijt}\right],$$

with $Z_{ijt} = (\text{constant}, dist_j, \mathcal{I}_{\tilde{\kappa},ijt})$ and $\mathbb{E}\left[r_{ijt} \mid \mathcal{I}_{\tilde{\kappa},ijt}\right] = \alpha_{jt}r_{iht} - e_{\tilde{\kappa}jt}$ denoting the predicted export revenue of an observed information type $\tilde{\kappa}$.

Since our approach relies on observing choice probabilities (market shares), we group firms with similar revenues $\alpha_{jt}r_{iht}$ into 20 deciles for each destination j in year t. The group-specific export probabilities are $s_{jct} = \frac{\sum_{i\in c} d_{ijt}}{N_c}$, where N_c is the number of firms in revenue decile c. We lose information by aggregating s_{ijt} to s_{gjt}. Consequently, some observations lead to moment inequalities that are too weak to reject any candidate parameter values. We exclude these data points from the estimation sample.[17]

We define the minimal information type as the observed type $\tilde{\kappa}$ in Assumption 4(i). Given that only about 11% of firms in the sample (chemical sector) are exporters, the estimated group-specific export probabilities s_{gjt} tend to suffer more from a downward aggregation bias. As a result, the min-side inequalities are much less informative than the max-side inequalities.

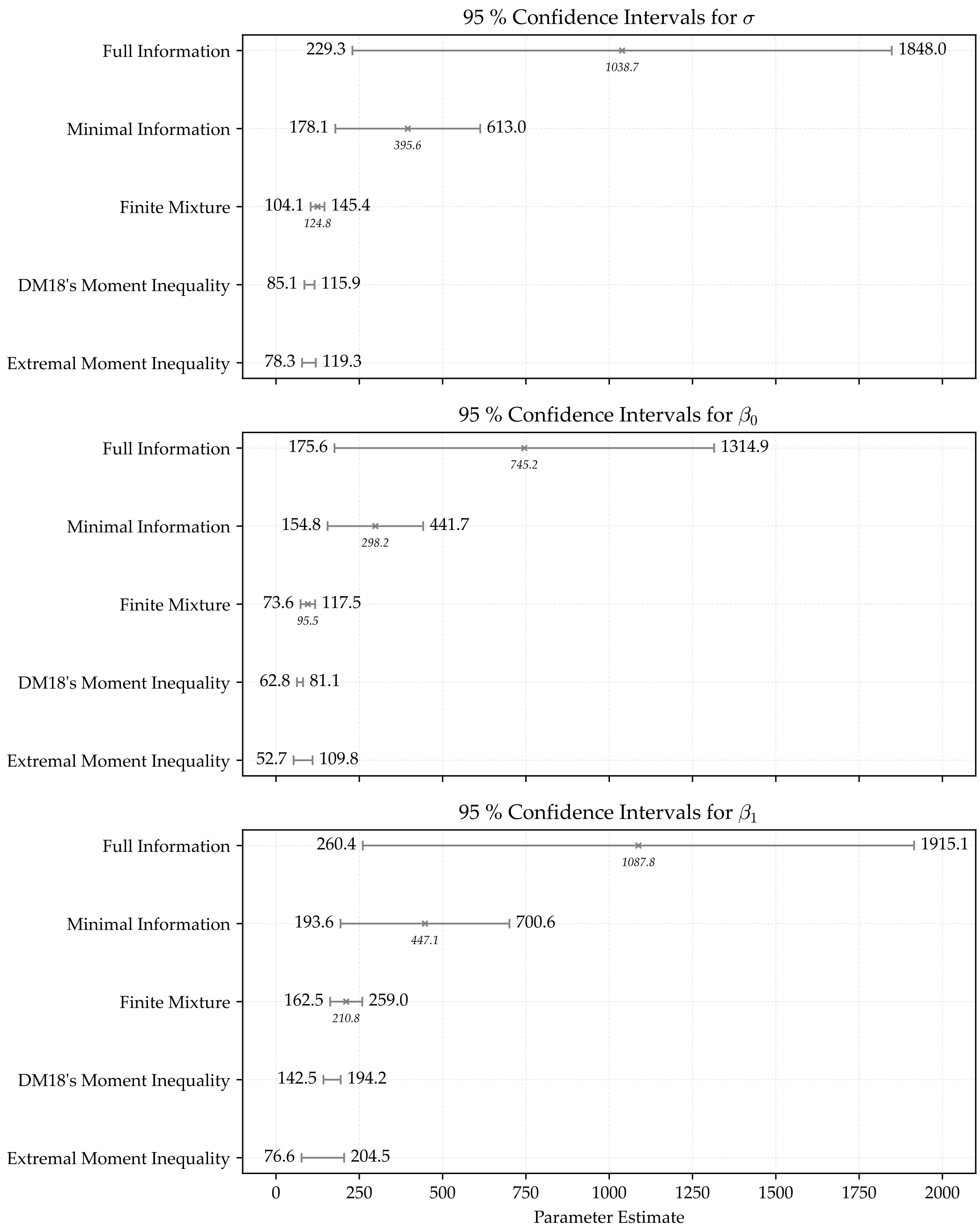
To address this issue, we modify the asymmetric selection thresholds in Assumptions 4(ii) and 4(iii). Specifically, after estimating the predicted export revenues $\hat{\mathbb{E}}\left[r_{ijt} \mid \mathcal{I}^m_{ijt}\right]$ for the minimal information type $\mathcal{I}^m_{ijt}$, we select extreme predictions below the 5-th percentile, i.e., $Z^S_{\min\, jt} = \mathbb{1}\left\{\hat{\mathbb{E}}\left[r_{ijt} \mid \mathcal{I}^m_{ijt}\right] \le q_{0.05}\right\}$, to construct the min-side moments. We use predictions above the 93-th percentile, i.e., $Z^S_{\max\, jt} = \mathbb{1}\left\{\hat{\mathbb{E}}\left[r_{ijt} \mid \mathcal{I}^m_{ijt}\right] \ge q_{0.93}\right\}$, to construct the max-side moments.

the predicted expectations vary across information types.

[17]Alternatively, we could specify an alternative parameter grid, but we prefer to evaluate the inequalities on the same grid as in DM18.

8.1.4 Results

Figure 2: 95% Confidence Intervals for the Chemical Sector



Notes. All estimates are reported in thousands of year 2000 US$, and their magnitudes scale proportionally with the parameter η, which is set to 5, following DM18. For each parameter σ, β_0, and β_1, we first plot 95% confidence intervals corresponding to three maximum likelihood estimators (MLE), with point estimates indicated by an 'x' at the center of each interval. The "Full Information" corresponds to the "Perfect Foresight" MLE in DM18, which assumes that firms' expectations coincide with realized outcomes, i.e., $\mathbb{E}[r_{ijt}|\mathcal{I}ijt] = rijt$. The "Minimal Information" replicates the specification in DM18 where firms form expectations using the minimal information set. The third row presents confidence intervals from our finite mixture model with four information types. Then, we plot the extreme points of the 95% confidence sets for the two moment inequality estimators. The "Finite Mixture" and the two moment inequality estimators are robust to unobserved information structure.

Table 5: Finite Mixture Estimates for Entry Decision in the Chemical Sector

Parameters	σ	β_0	β_1	ϕ_A	ϕ_B	ϕ_C	ϕ_D
	124.8	95.5	210.8	0.15	1.17E-12	0.56	0.29
	(11)	(11)	(25)	(0.048)	(0.092)	(0.078)	(–)

Notes. All the preference estimates (σ, β_0, β_1) are reported in thousands of year 2000 US$, and their values scale proportionally with η, which is set equal to 5, as in DM18. The finite mixture model consists of four information types $\kappa = A, B, C, D$ and the estimated share of the information type D is calculated as $\phi_D = 1 - \phi_A - \phi_B - \phi_C$. Bootstrap standard errors are computed as in DM18 and reported in parentheses.

Figure 2 compares parameter estimates for the chemical sector for 5 specifications. First, as in DM18, we show results for two naive plug-in approaches, i.e. assuming that all exporters have full information or minimal information so that there is no heterogeneity in information sets. Next, we replicate the moment inequality of DM18 based on the minimal information, which finds small estimates and, thus, a smaller fixed cost of exporting than naive approaches. Finally, the figure compares these results with our moment inequality and finite mixture approach results, where we find our results perform just as well as DM18.

Table 5 presents the estimation results from the finite mixture model in the chemical sector. From the estimated fractions ϕ_κ, we observe a significant share of information type A, the minimal information type, confirming the findings in DM18. 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 close to zero and not statistically significant. Intuitively, information about the number of total exporters might be easier to obtain for a firm than the revenue shifter. However, there is also a substantial proportion of firms that use all information variables ($\hat{\phi}_D = 0.29$).

8.2 Consumer Valuation of Fuel Costs

GRV18 studies the valuation of automobile fuel costs in the EU market. The authors use a demand model with rich consumer heterogeneity to obtain estimates of consumers' valuation of fuel costs. Specifically, they allow preference heterogeneity using an empirical mileage distribution and random coefficients. 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), that looks at identifying the responsiveness of purchases of energy-consuming durables to their energy expenses. This literature has relied on plug-in approaches. 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.

We extend GRV18 and investigate the information variables consumers use to form expectations of fuel costs. We recast the problem in our framework and estimate a finite mixture of observable information types. This allows us to obtain estimates of consumers' valuation of their expected fuel costs consistent with a much richer information structure and to 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 (defined as year-country) t. For simplicity, we omit the market subscript t. Her decision utility is defined as

$$u^e_{ijk} = x_{jk}\beta^x_i - \alpha_i(p_{jk} + \gamma G_{ijk}) + \xi_{jk} + \epsilon_{ijk},$$

where each car is defined based on its model j and engine type k, x_{jk} is a (row) vector of observed car characteristics, p_{jk} is the price, G_{ijk} is consumer i's expected fuel cost, ξ_{jk} is an unobserved product attribute and ϵ_{ijk} is the idiosyncratic valuation for the car, modeled as an EVT1 random variable. The vector β^x_i includes 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

$$G_{ijk} = \rho\beta^m_i e_{jk} f_k,$$

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^m_i is consumer i's annual mileage, e_{jk} is the fuel consumption of the car and f_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}.

GRV18 assumes that consumers have full information about 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 consider the observed fuel efficiency e_{jk} and fuel prices f_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 $\mathcal{W}_A = \{e_{jk}\}$, $\mathcal{W}_B = \{f_k\}$, and $\mathcal{W}_C = \{e_{jk}, f_k\}$, respectively. We denote the product $e_{jk}f_k$ as the uncertain attribute g_{jk} in our model. The expectation of each information type is $g^e_{\kappa jk} = \mathbb{E}[e_{jk}f_k|\mathcal{W}_\kappa]$. The information type C corresponds to the perfect foresight type that uses both fuel efficiency and fuel price to predict fuel costs.

The market share for model j with engine k equals

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

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. 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 GRV18, we restrict the random coefficients to be $\beta_i^x = \bar{\beta}^x + \Sigma^x \nu_i^x$, 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 inversely proportional to the 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. 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.

We apply a random forest prediction to approximate the unobserved shape of the conditional expectation operator $g^e_{\kappa jk} = \mathbb{E}[e_{jk}f_k|\mathcal{W}_\kappa]$. This allows us to capture the non-linear relationship between the information variables and the expected fuel costs, as displayed in Appendix Figure A4.

We first estimate the mixture model in Equation (21) without random coefficients, i.e., $\beta^x_i = \bar{\beta}^x$ 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). 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 6 presents the estimation results. The first three columns replicate results from Table 3 in GRV18 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 nearly zero. We find that more than half of the consumers use only fuel efficiency to predict future fuel costs, while the remainder rely solely on fuel price. Finally, our mixture model estimates $\hat{\gamma} = 0.23$ much lower than the 0.91 estimated under the full information model. This suggests that accounting for consumers' unobserved information about fuel costs is important in this setting. Our 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 6: Parameter Estimates for Alternative Demand Models

	Table 3 in Grigolon et al. (2018)			Finite Mixture		
	Logit	RC Logit I	RC Logit II	Logit	RC Logit I	RC Logit II
	(1)	(2)	(3)	(4)	(5)	(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)	(–)	(–)	(–)

Notes. This table reproduces estimates from GRV18 in columns (1)-(3) and applies the finite mixture model with three information types in columns (4)-(6). Standard errors are reported in parentheses.

9 Conclusion

This paper introduces two novel econometric approaches to estimating discrete choice models in the presence of unobserved heterogeneous information sets. We illustrate the significance of accounting for unobserved information heterogeneity through empirical applications and simulations. Our analysis of firm export decisions confirms that our methods successfully find estimates similar to previous approaches but under very different assumptions. In the context of consumer vehicle purchases, our results reject the commonly assumed perfect foresight about fuel costs and reveal substantial heterogeneity in how consumers form expectations.

Our methodological contributions extend existing work on discrete choice under uncertainty by introducing new tools that can be readily implemented in empirical research. The finite mixture model offers a tractable solution for recovering informational heterogeneity using standard choice data. At the same time, the moment inequality approach provides a more robust alternative in settings where the researcher cannot specify all information variables. Given the increasing relevance of policy interventions that aim to influence choice through information provision, our framework is valuable for assessing how informational heterogeneity affects counterfactual predictions.

Future research may explore extensions of our methods to dynamic decision-making settings and contexts with dynamic learning. Additionally, further work is needed to understand how these approaches can be applied in high-dimensional settings where the set of potential information variables is large. By bridging the gap between observed choices and unobserved expectations, our work advances the empirical toolkit available to researchers studying economic decision-making under uncertainty.

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.

Agarwal, Nikhil and Paulo J Somaini (2022) "Demand analysis under latent choice constraints,"Technical report, National Bureau of Economic Research.

Ahn, Hyungtaik and Charles F Manski (1993) "Distribution theory for the analysis of binary choice under uncertainty with nonparametric estimation of expectations," *Journal of Econometrics*, 56 (3), 291–321, https://doi.org/10.1016/0304-4076(93)90123-M.

Allcott, Hunt and Christopher Knittel (2019) "Are Consumers Poorly Informed about Fuel Economy? Evidence from Two Experiments," *American Economic Journal: Economic Policy*, 11 (1), 1–37, 10.1257/pol.20170019.

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.

Barahona, Nano, Cristóbal Otero, and Sebastián Otero (2023) "Equilibrium Effects of Food Labeling Policies," *Econometrica*, 91 (3), 839–868, https://doi.org/10.3982/ECTA19603.

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.

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 and James J. Heckman (2007) “Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling,” *Labour Economics*, 14 (6), 870–893, https://doi.org/10.1016/j.labeco.2007.06.002, Education and Risk.

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 (2023) "Estimating demand for differentiated products with zeroes in market share data," *Quantitative Economics*, 14 (2), 381–418, https://doi.org/10.3982/QE1593.

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.

Manski, Charles F. (1993) *2. Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?*, 43–60, Chicago: University of Chicago Press, doi:10.7208/9780226110646-005.

——— (2004) "Measuring Expectations," *Econometrica*, 72 (5), 1329–1376, http://www.jstor.org/stable/3598892.

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.

Nevo, Aviv, John L Turner, and Jonathan W Williams (2016) "Usage-based pricing and demand for residential broadband," *Econometrica*, 84 (2), 411–443.

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`.

Appendices

A Proof of Partial Identification

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 depend on these simplifications made for illustrative reasons. DM18 presents a similar proof for the setting with micro-choice data.

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 us 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):

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

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 $f^y(g_{ijt}|\mathcal{K}_{jt}, g^e_{ijt})$ and $f^y(g^e_{ijt}|\mathcal{K}_{jt})$.

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

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

Second, by Bayes' rule, we have

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

where $f(\mathcal{K}_{jt}|g^e_{ijt})$ is determined by $\mathcal{K}_{jt}|g^e_{ijt} \sim \mathcal{N}(g^e_{ijt} + \mu_v, \sigma^2_v)$, $f(\mathcal{K}_{jt})$ is obtained with $\int f(\mathcal{K}_{jt}|g^e_{ijt}) dF(g^e_{ijt})$ where $F(g^e_{ijt})$ is the CDF of $\mathcal{N}(\mu_{g^e}, \sigma^2_{g^e})$, and $f(g^e_{ijt})$ 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 $(g^e_{ijt}, v_{ijt}, e_{ijt})$ 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

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

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:[18]

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

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 be solved 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.[19]

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*

[18] 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, $\text{Cov}(g^e, e) = 0$.

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

$\{s_{\kappa 0t}\}$ *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., $\frac{s_{jt}}{s_{0t}} = \frac{\sum_{\kappa=1}^{K_3} \phi_\kappa s_{\kappa jt}(\delta_t, g^e_{\kappa t}; \theta)}{\sum_{\kappa=1}^{K_3} \phi_\kappa s_{\kappa 0t}(\delta_t, g^e_{\kappa t}; \theta)}$. 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:

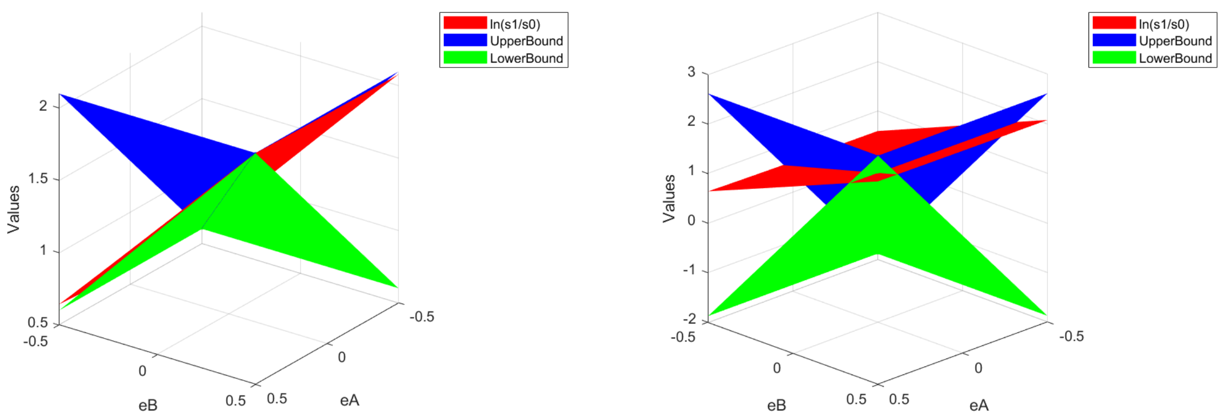
$$\min_{\kappa}\{-\gamma e_{\kappa jt}\} + \gamma g_{jt} + X_{jt}\beta + \xi_{jt} \leq \log \frac{s_{jt}}{s_{0t}} \leq \max_{\kappa}\{-\gamma e_{\kappa jt}\} + \gamma g_{jt} + X_{jt}\beta + \xi_{jt},$$

which gives the moment inequalities in Equation (13).

C Additional Figures

Figure A1: Graphical Illustration of Inequalities

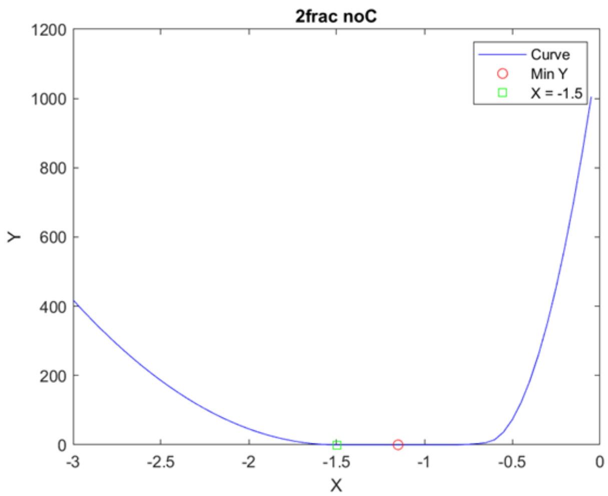
(a) Under True Value of $\gamma = -1.5$ (b) Under False Value of $\gamma = -4.5$



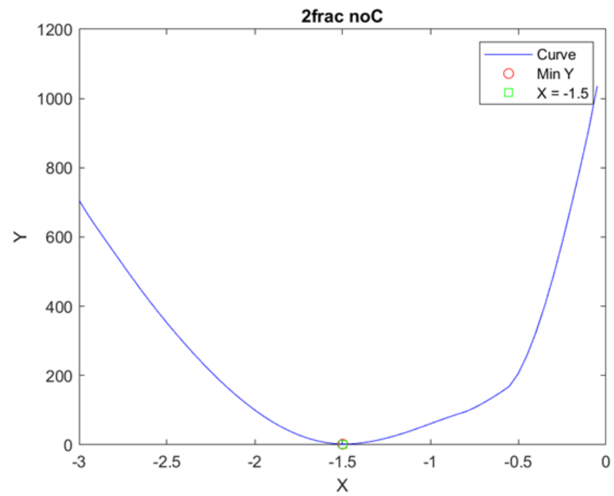
Notes. This figure illustrates the power of our inequalities (14) in identifying the correct values of the parameter γ. The right-hand side of (14) is plotted in blue, while the left-hand side is in green. The observed log market share ratio is plotted in red. Given the true parameter value $\gamma = -1.5$, plot A1a shows that the max side in blue always bounds the market share ratio from above while the min side in green bounds the data from below. However, given a wrong parameter value $\gamma = -4.5$, our inequalities are violated, which will then rule out such wrong parameter value from the identified set, as depicted in plot A1b.

Figure A2: Comparison of GMM Objective Function with vs. without Chamberlain IVs

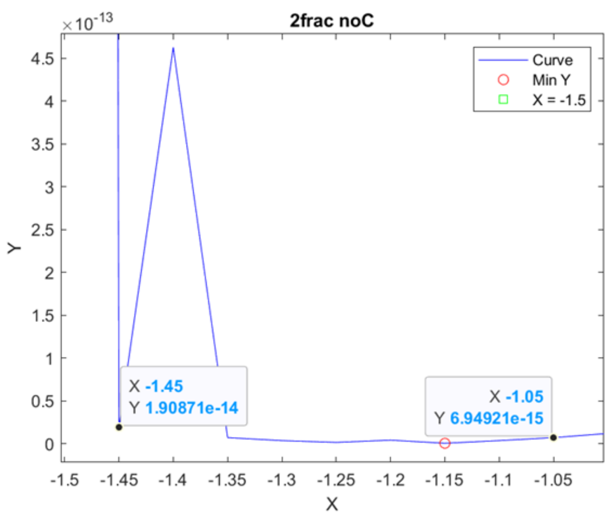
(a) Regular IVs: Gobal View



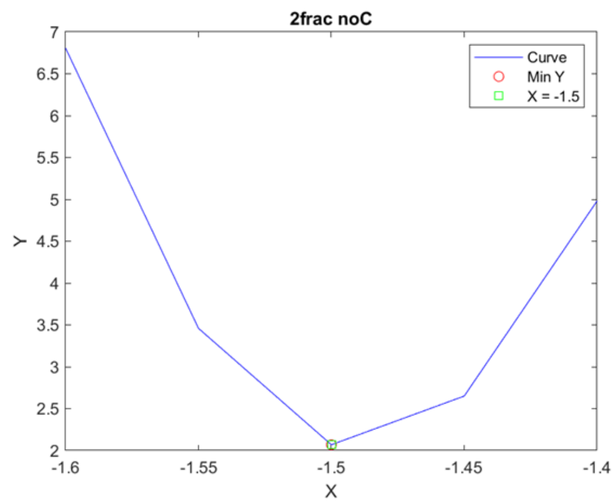
(c) Chamberlain IVs: Gobal View



(b) Regular IVs: Local View

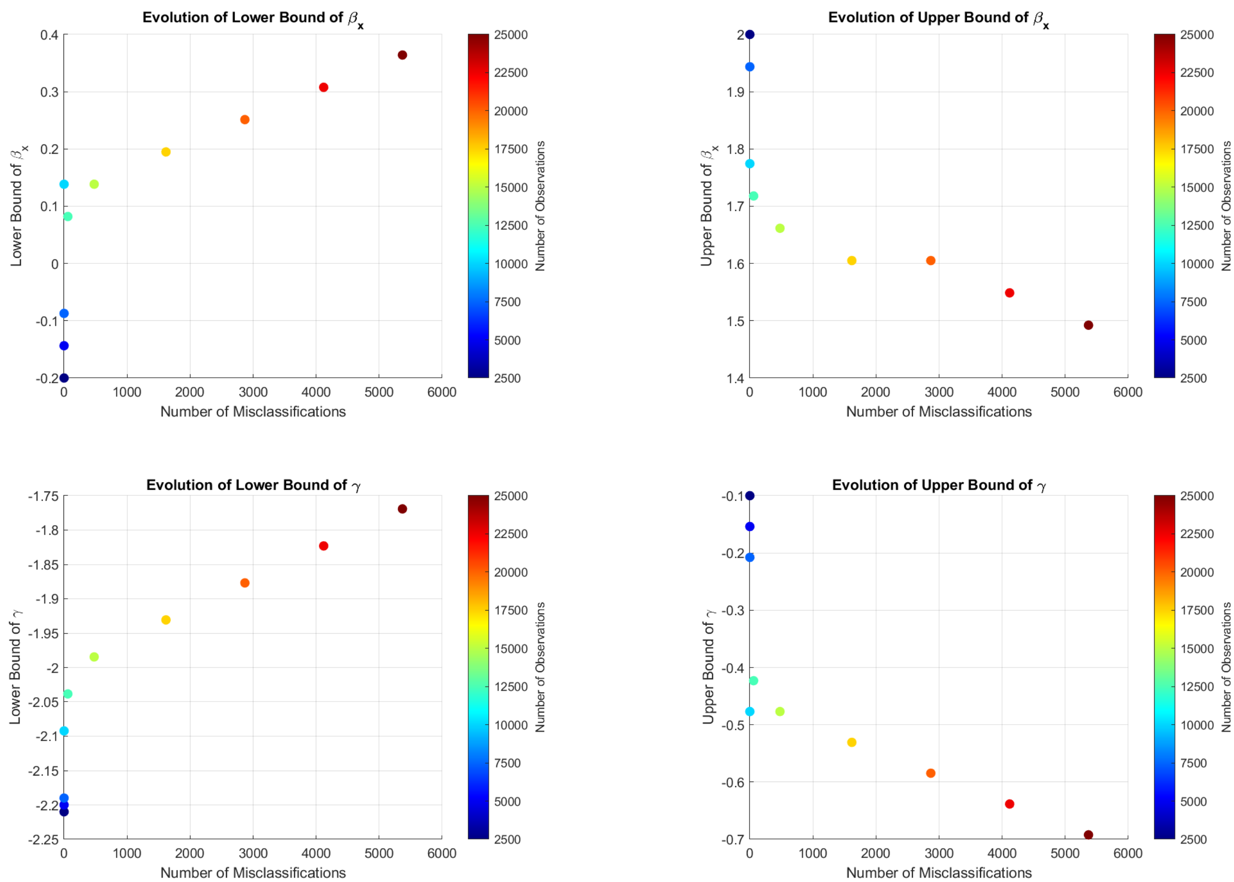


(d) Chamberlain IVs: Local View



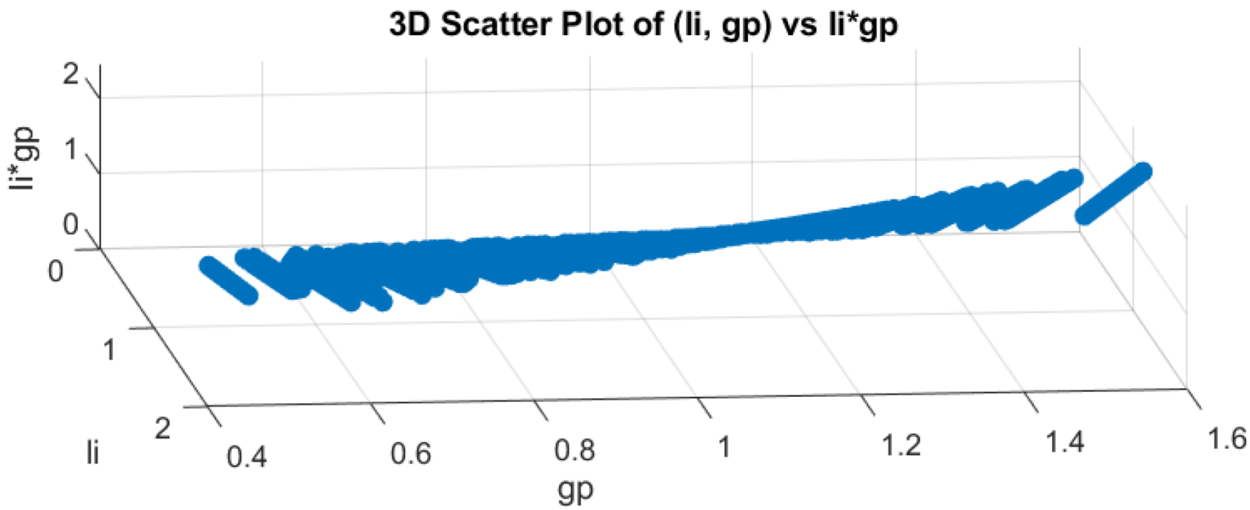
Notes. This figure displays the GMM objective function using two types of instruments: standard IVs in panels (a) and (b) on the left, and Chamberlain IVs in panels (c) and (d) on the right. The top panels (a, c) illustrate the global shape of the objective function, while the bottom panels (b, d) focus on the region where the function attains its minimum. The horizontal axis shows the estimated parameter γ; the vertical axis shows the value of the objective function. The true parameter value, $\gamma = -1.5$, is marked by the green box, and the red circle indicates the objective function's minimum. We observe that the regular instruments lead to a problematic objective function with multiple local minima. However, the GMM objective function becomes well-behaved when using Chamberlain's instruments, further supports the effectiveness of Chamberlain's instruments.

Figure A3: Estimated bounds for β_x and γ across different selection thresholds ($\tilde{\kappa} = A$)



Notes. This figure plots the estimated bounds for each parameter β_x and γ across different thresholds. The y-axis corresponds to the value of the estimated bound. The x-axis corresponds to the number of misclassification in the estimation sample. The color indicates the sample size. We observed that, as the sample size increases, the confidence interval becomes more precise (the lower bound increases while the upper bound decreases).

Figure A4: Non-linear Relationship between Information Variables and Fuel Costs



Notes. This figure plots the distribution of the fuel costs (g_{jkt}, i.e., "li*gp" on the z-axis) as a function of two information variables–fuel prices (f_{kt}, i.e., "gp" on the x-axis) and fuel efficiency (e_{jkt}, i.e., "li" on the y-axis). The fuel costs are defined as $g_{jkt} := e_{jkt} \cdot f_{kt}$.