

# Estimating Choice Models with Unobserved Expectations over Attributes

Mathias Reynaert

Wenxuan Xu

Hanlin Zhao \*

October 12, 2026

## Abstract

Agents often choose based on expectations about uncertain attributes, yet researchers rarely observe the information behind those expectations. We develop two methods to estimate discrete choice models with heterogeneous, unobserved information sets. Our point-identification approach uses repeated realizations of the uncertain attribute to estimate its statistical process with observed predictors and recoverable latent factors, then represents conditional expectations as a finite mixture over these components. A complementary approach set-identifies preferences using partial information about expectations. Both methods apply to individual- and market-level data. Applications to firms' export decisions and consumers' automobile purchases show that our method corrects for substantial bias in preference estimates that arises when unobserved information heterogeneity is ignored.

*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: National University of Singapore, and Zhao: Toulouse School of Economics, University of Toulouse Capitole. We thank Christian Bontemps, Zach Brown, Federico Bugni, Chris Conlon, Michael Dickstein, Gaston Illanes, Arnaud Maurel and seminar participants at Northwestern University, Toulouse School of Economics, Vienna University, University of Helsinki, APIOC 2024, IIOC 2025 and EARIE 2025 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 economic choices depend on expectations about uncertain attributes. Consumers anticipate future fuel costs when purchasing vehicles (Hausman, 1979), individuals form expectations about their health risk when selecting insurance (Einav et al., 2010), and students forecast future earnings when choosing educational pathways (Arcidiacono et al., 2020). In each setting, observed choices reflect both agents' preferences over attributes and their expectations about those attributes. A low willingness to pay for fuel efficiency, for example, can reflect either a low valuation of future operating-cost savings or an expectation that those savings will be small. Distinguishing these explanations matters for estimating preferences and predicting responses to changes in economic conditions and information.

Even when researchers observe realized attributes, they generally do not observe the expectations that entered agents' choices or the information used to form them. Agents may also rely on different information when predicting the same attribute. A common empirical approach specifies a prediction rule and substitutes the resulting expectations into a choice model. Substituting realized attributes for expectations, for example, imposes perfect foresight. When the implied expectations differ from those entering agents' decisions, preference estimates can be biased (Manski, 1993, 2004; Cunha and Heckman, 2007). Without additional restrictions on how expectations vary across agents, observed choices cannot generally identify preferences separately from information heterogeneity.

This paper develops two complementary methods for estimating discrete-choice models with heterogeneous and unobserved information sets. Our main approach constructs a finite collection of prediction rules and jointly estimates preferences and the share of agents using each rule. We establish identification conditions and develop estimators using either individual choices or aggregate market shares. The finite-mixture framework accommodates multinomial choice, random coefficients, and endogenous characteristics, allowing information heterogeneity to be incorporated into workhorse empirical models, including the demand systems of Berry et al. (1995) (BLP). A complementary moment-inequality approach bounds preferences without recovering or enumerating the complete collection of prediction rules. It uses a prediction rule associated with one information type and observations on which that rule produces the largest or smallest prediction error across types. Both methods use data on realized attributes and choices without requiring direct measures of agents' expectations.

The finite-mixture estimator proceeds in two stages. First, we use repeated realizations

of the uncertain attribute to estimate its statistical process. Its predictive components can include observed variables and latent factors or fixed effects recovered from the repeated data. Conditioning on different combinations of these components generates a finite collection of rational-expectations prediction rules (Ahn and Manski, 1993), each defining an information type. We require each agent’s vector of conditional expectations across alternatives to coincide with the predictions generated by one of these rules. In the second stage, these predictions enter a finite-mixture choice model, whose weights measure the shares of information types. With individual choice data, we use maximum likelihood; with aggregate shares, we use generalized method of moments based on BLP share inversion.

Identification exploits differences across information types in how predictive variables affect expected attributes and hence choice probabilities. As these variables vary across alternatives and markets, the resulting choice patterns provide information about preferences and the prevalence of each type. Given the recovered prediction functions, we establish explicit rank conditions for local joint identification of preferences and information-type shares in the baseline model. We extend the identification argument to random coefficients and endogenous characteristics using the results of Berry and Haile (2014) and Wang (2023). This extension maintains a common distribution of tastes across information types and imposes instrument and support conditions that permit separate recovery of preference heterogeneity and information-type shares.

The estimated model supports counterfactuals that change either the economic environment or the distribution of information types. Researchers can change prices or other attributes while holding the information distribution fixed, or change the prevalence of information types while holding preferences fixed. The type-specific choice probabilities translate these changes into predicted behavior. Researchers can specify the counterfactual type shares using a model of information acquisition or evidence from information interventions, such as fuel-economy disclosure or food labeling (Allcott and Knittel, 2019; Barahona et al., 2023).

The moment-inequality approach applies when recovering or enumerating the complete collection of prediction rules is impractical. It requires the researcher to observe the information set used by one type present in the population and to construct that type’s predictions. Under the finite-mixture logit structure, the log ratio of a product’s share to the outside-option share lies between the smallest and largest systematic utility indices across information types. Selecting observations on which this type’s prediction error is maximal or minimal across all types allows us to construct the corresponding bound using its prediction, which is in a similar spirit to D’Haultfœuille et al. (2018). These bounds

yield implementable moment inequalities for preferences and associated confidence sets. Selection entails a validity–power trade-off: tighter thresholds can make the extremal-selection restriction more credible but reduce the effective sample size and may lower statistical power. Our simulations illustrate the sensitivity of the resulting confidence sets to this choice.

Two applications demonstrate how the framework can be used with existing empirical models and data. First, we revisit Dickstein and Morales (2018) (DM18 hereafter), who study Chilean firms’ export decisions under uncertainty about export revenues. Using their construction of potential revenues, we estimate a finite mixture of prediction rules that share their minimal information set and differ in the additional variables used to forecast revenues. The finite-mixture confidence intervals and the confidence sets from our extremal inequalities overlap the corresponding DM18 confidence sets for exporting fixed costs. The estimated information-type shares indicate substantial heterogeneity: in the chemical sector, approximately 15% of firms use only the minimal set, 56% add only lagged exporter counts, and 29% use all five candidate variables.

Second, we incorporate information heterogeneity into the automobile demand model of Grigolon et al. (2018) (GRV18 hereafter). This application uses aggregate market shares and allows for heterogeneity in tastes and mileage, endogenous vehicle prices, and unobserved product characteristics. Consumers form fuel-cost predictions using fuel consumption, fuel prices, or both. In the preferred random-coefficient specification, the estimated shares are 51% for the fuel-consumption rule, 39% for the fuel-price rule, and 10% for the rule using both, the last of which corresponds to perfect foresight in this application. Allowing for these heterogeneous predictions substantially lowers the inferred valuation of future fuel costs. The estimated internalization parameter, which measures the weight on discounted expected fuel costs relative to the purchase price, falls from 0.91 in GRV18 to 0.22.

Accounting for information heterogeneity also changes the relative effectiveness of tax instruments. In the policy benchmark of GRV18, where consumers are estimated to nearly fully internalize expected fuel costs, a fuel tax reduces aggregate fuel consumption by more than a revenue-equivalent product tax. We find that this ranking reverses once consumer information heterogeneity is incorporated. Even under perfect foresight, our lower estimated degree of fuel-cost internalization substantially weakens the response to a fuel tax: a €0.50-per-liter fuel tax reduces annual fuel consumption by 3.9%, compared with 14.4% under the revenue-equivalent product tax. Allowing consumers to use heterogeneous prediction rules further widens this gap, particularly because some consumers form expectations without conditioning on fuel prices. In this case, the fuel tax reduces fuel

consumption by only 2.6%, whereas the product tax reduces it by 18.1%.

Our approach builds on the use of realized outcomes to study agents' information under rational expectations. Cunha et al. (2005) (CHN05 hereafter) and Cunha and Heckman (2007) use outcome processes and choice behavior to distinguish predictable components of outcomes from uncertainty at the time of choice, allowing for latent factors whose realizations differ across individuals. Related frameworks have been applied to educational choices and skill formation (Aucejo and James, 2021; Houmark et al., 2024). Our finite mixture extends this approach by allowing agents to use different combinations of the predictive components and estimating the share using each combination. The common information structure across agents in the existing literature is nested in our approach as a mixture concentrated on one type.

A second strand of research derives preference bounds while allowing part of agents' information to remain unobserved. DM18 construct inequalities from individual export decisions using a researcher-observed information set contained in every agent's information set, with related approaches extended in Dickstein et al. (2024) and Porcher et al. (2024). Our inequalities replace this common-information requirement with a prediction rule used by one type and selection events on which its prediction errors are extremal. More broadly, Bergemann et al. (2022) study counterfactuals with latent information, and Gualdani and Sinha (2024) characterize sharp preference and counterfactual bounds in discrete-choice models with heterogeneous information structures. Our finite-mixture approach restricts predictions to a recoverable finite collection, allowing us to estimate preferences and information-type shares within BLP demand systems.

Finally, our paper relates to work on consumer search and limited consideration, which studies unobserved differences in the information entering choice (Abaluck and Adams-Prassl, 2021; Abaluck et al., 2026). We model information through conditional expectations of uncertain attributes. Other empirical approaches study expectations and information formation using reported beliefs, experiments, or observed quality signals (Arcidiacono et al., 2020; Brown and Jeon, 2024; Vatter, 2025). Our framework relies on repeated realizations of uncertain attributes and choice data to estimate preferences and information use when direct measures of expectations are unavailable.

Section 2 presents the choice model, and Section 3 explains the identification challenge. Section 4 develops the finite-mixture approach, and Section 5 presents the moment inequalities. Section 6 reports the simulations, and Section 7 presents the empirical applications.

## 2 Choice Model

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

$$u_{ijt} \equiv X'_{ijt}\beta + \gamma g_{ijt} + \epsilon_{ijt}, \quad (1)$$

where  $X_{ijt} \in \mathbb{R}^{|X|}$  is a vector of observed attributes,  $g_{ijt} \in \mathbb{R}$  denotes the realized value of an attribute that is uncertain to individuals at the time of decision,  $\epsilon_{ijt} \in \mathbb{R}$  is an idiosyncratic taste shock that is i.i.d. across options  $j$ , conditional on  $(i, t)$ . The researcher observes  $(X_{ijt}, g_{ijt})$ .<sup>1</sup> The outside option utility is normalized to  $u_{i0t} = \epsilon_{i0t}$ . The vector of preference parameters of interest is  $\theta = (\beta', \gamma)'$ . We begin by presenting the model under homogeneous preferences and exogenous characteristics. Section ?? extends the framework to allow for preference heterogeneity and endogenous attributes.

Let  $\mathcal{I}_{it}$  denote individual  $i$ 's information set in market  $t$  which includes known attributes, the primitives, and signals about  $g_{ijt}$ . We assume that individuals use a subset of this information to form predictions about the uncertain attribute  $g_{ijt}$ . Let  $\mathcal{W}_{it} \subseteq \mathcal{I}_{it}$  denote the collection of information variables used by individual  $i$  to predict  $g_{ijt}$ . We impose rational expectations, so that

$$g_{ijt}^e \equiv \mathbb{E}[g_{ijt} \mid \mathcal{W}_{it}].$$

Expected utility is given by

$$u_{ijt}^e \equiv X'_{ijt}\beta + \gamma g_{ijt}^e + \epsilon_{ijt}, \quad (2)$$

Individuals observe  $g_{ijt}^e$  when making decisions, whereas the researcher does not. The expected utility of the outside good is normalized to  $u_{i0t}^e = u_{i0t} = \epsilon_{i0t}$ . Individuals choose the option that maximizes expected utility:

$$d_{ijt} \equiv \mathbb{1} \left\{ u_{ijt}^e \geq \max_{j' \in \mathcal{J}} u_{ij't}^e \right\},$$

with  $d_{ijt} = 1$  if  $i$  chooses option  $j$  in market  $t$  and  $d_{ijt} = 0$  otherwise. We formalize these

---

<sup>1</sup>As in DM18, the researcher observes the realization of  $g_{ijt}$  for all  $j$ . This differs from settings with selection (as in CHN05), where  $g_{ijt}$  is only observed for the chosen option.

modeling assumptions as follows:

**Assumption 1** (Choice under uncertainty). *At the time of choice, individuals are uncertain only about  $g_{ijt}$  and form rational expectations conditional on  $\mathcal{W}_{it}$ . They choose the option that maximizes expected utility given in (2). Taste shocks  $\epsilon_{ijt}$  are i.i.d. across  $j$  conditional on  $(i, t)$  and are independent of  $(X_{ijt}, \mathcal{W}_{it})$ .*

We define the information variables comprising  $\mathcal{W}_{it}$  as follows:

**Definition 1.** *A candidate variable  $w$  is:*

- (a) **Predictive** if  $\mathbb{E}[g \mid w] \neq \mathbb{E}[g]$ ;
- (b) **Used** if  $w \in \mathcal{W}_{it}$ , so that  $g_{ijt}^e$  depends on  $w$ .

*An information variable is a candidate variable that is both predictive and used.*

In practice, information variables may include lagged realizations of the uncertain attribute, expert forecasts, or publicly available statistics related to  $g_{ijt}$ . As in revealed-preference settings such as CHN05, we do not distinguish between variables that are observed but not used by agents and variables the individual does not observe; both are treated as excluded from  $\mathcal{W}_{it}$ .

We formalize the prediction information set as

$$\mathcal{W}_{it} = \{w_{m,ijt} : m \in \mathcal{M}_i, j \in \mathcal{J}\},$$

where  $\mathcal{M}_i$  denotes the index set of information variables used by individual  $i$ , and each  $w_{m,ijt}$  may vary across options.<sup>2</sup>

**Remark 1.** *Our framework allows observed attributes  $X_{ijt}$  to enter  $\mathcal{W}_{it}$  and thus be used in forming predictions about  $g_{ijt}$ . For simplicity, we abstract from this going forward and assume that agents do not use  $X_{ijt}$  to predict  $g_{ijt}$ . This restriction allows us to distinguish between variables that shift utility directly and variables that operate exclusively through agents' information sets.*

Information heterogeneity arises in two ways. First, individuals may rely on different subsets of information variables, so that  $\mathcal{M}_i \neq \mathcal{M}_{i'}$ . Second, even when using the same information variables, realizations may differ across individuals. This structure follows

---

<sup>2</sup>For simplicity, we assume that  $\mathcal{M}_i$  is stable across markets. The framework can be extended to allow for market-specific information sets.

DM18 in allowing both heterogeneity in information sets and heterogeneity in realizations. In contrast, CHN05 allows heterogeneous realizations but imposes a common information set.

**Assumption 2** (Information heterogeneity). *The index set  $\mathcal{M}_i$  is individual-specific and may not be common across individuals. Realizations of the information variables may also differ across individuals.*

The researcher observes a random sample of  $T$  markets. In each market  $t$ , the researcher observes either individual choices  $d_t = \{d_{ijt}\}_{i \in \mathcal{N}_t, j \in \mathcal{J}}$  or aggregate market shares  $\{s_{jt}\}_{j \in \mathcal{J}}$ . Let  $X_t = \{X_{ijt}\}_{i \in \mathcal{N}_t, j \in \mathcal{J}}$  denote the observed attributes and let  $g_t = \{g_{ijt}\}_{i \in \mathcal{N}_t, j \in \mathcal{J}}$  denote the realizations of the uncertain attribute. Based on the data  $\{d_t \text{ or } s_t, X_t, g_t\}$ , the researcher's objective is to identify the preference parameters  $\theta$  and the distribution of information-variable index sets  $\mathcal{M}_i$  in the population. We close this section with two illustrative examples that serve as our empirical applications.

**Example 1** (Exporting Firms). The researcher observes individual-level data  $\{d_t, X_t, g_t\}_{t \in \mathcal{T}}$ . This corresponds to the setting in DM18, where the researcher observes the export decision  $d_{ijt}$  of each firm  $i$  to destination country  $j$  in year-sector  $t$ .

The utility function  $u_{ijt}$  in Equation (1) corresponds to exporter  $i$ 's profit from exporting to  $j$ , which depends on observed attributes  $X_{ijt} = X_{jt}$  such as bilateral distance. Profits also depend on export revenue  $g_{ijt}$ , which is uncertain at the time of the export decision. Each exporter therefore forms a prediction  $g_{ijt}^e = \mathbb{E}[g_{ijt} \mid \mathcal{W}_{it}]$  based on its information set.

Assuming that the idiosyncratic error term  $\epsilon_{ijt}$  follows a type-I extreme value distribution, the conditional choice probability given  $\mathcal{W}_{it}$ , obtained by integrating over  $\epsilon$ , is

$$\Pr(d_{ijt} = 1 \mid X_t, \mathcal{W}_{it}; \theta) = \frac{\exp(X'_{jt}\beta + \gamma g_{ijt}^e)}{1 + \sum_{j'=1}^J \exp(X'_{j't}\beta + \gamma g_{ij't}^e)}.$$

**Example 2** (Automobile Purchase). The researcher observes market-level data  $\{s_t, X_t, g_t\}_{t \in \mathcal{T}}$ , as in Berry (1994) and Berry et al. (1995). For example, Grigolon et al. (2018) use market-level data to estimate consumers' valuation of fuel costs when purchasing automobiles.

The utility function  $u_{ijt}$  in Equation (1) corresponds to consumer  $i$ 's utility from purchasing vehicle  $j$  in year-country  $t$ . Utility depends on observed attributes  $X_{jt}$  such as the vehicle price  $p_{jt}$ , and on fuel costs  $g_{jt}$  that are uncertain at the time of purchase. Each consumer forms a prediction  $g_{ijt}^e = \mathbb{E}[g_{jt} \mid \mathcal{W}_{it}]$  based on her information set.

Assuming that  $\epsilon_{ijt}$  follows a type-I extreme value distribution, the market share of product  $j$  in market  $t$  integrates the logit choice probability over the cross-sectional

distribution of consumers' predictions. Let  $g_i^e = (g_{i1t}^e, \dots, g_{iJt}^e)$  denote consumer  $i$ 's vector of predicted fuel costs across products in market  $t$ , and let  $F_t(g_i^e)$  denote the cross-sectional distribution of this prediction vector. The predicted market share is

$$s_{jt}(X_t; \theta) = \int \frac{\exp(X'_{jt}\beta + \gamma g_{ijt}^e)}{1 + \sum_{j'=1}^J \exp(X'_{j't}\beta + \gamma g_{ij't}^e)} dF_t(g_i^e).$$

The researcher does not observe consumers' predictions  $g_{ijt}^e$ , but may observe individual-specific realizations  $g_{ijt}$  in datasets with individual-level outcomes. In the market-level automobile application considered here, the realized per-mile fuel cost is instead common across consumers,  $g_{ijt} = g_{jt}$ , and is constructed from the vehicle's fuel-consumption rate and the relevant fuel price. Individual mileage enters expected total expenditure separately.

### 3 Identification Challenges

This section explains why unobserved information can prevent joint identification of preferences and predictions, and why treating realized attributes as endogenous does not by itself resolve this problem.

Recall that agents' choices depend on expected attributes  $g_{ijt}^e \equiv \mathbb{E}[g_{ijt} \mid \mathcal{W}_{it}]$ , where  $\mathcal{W}_{it}$  denotes the individual-specific information set used to form predictions. The researcher observes realized attributes  $g_{ijt}$  but neither  $\mathcal{W}_{it}$  nor the mapping from  $\mathcal{W}_{it}$  to predictions  $g_{ijt}^e$ . Under Assumption 2, individuals facing identical realizations  $(X_{ijt}, g_{ijt})$  may receive different signals or use different subsets of those signals. Their rational predictions, and hence their choice probabilities, can therefore differ.

A common response is to treat the realized attribute  $g_{ijt}$  as endogenous and consider an instrumental-variable (IV) strategy. An instrument that shifts realized outcomes, however, need not shift the predictions that enter utility. Valid IV estimation requires restrictions linking the instrument to agents' predictions and prediction errors, as well as to taste shocks. These restrictions do not follow from exclusion from taste shocks alone when the prediction information sets are unobserved.<sup>3</sup>

A common "plug-in" approach replaces the unobserved prediction  $g_{ijt}^e$  in (2) with the realized attribute  $g_{ijt}$ , effectively imposing perfect foresight. Under imperfect information,

---

<sup>3</sup>Related identification strategies exploit variation in publicly available information: Brown (2019) uses the introduction of a price-transparency website, while Brown et al. (2023) use discontinuities in displayed provider ratings to distinguish preferences from information frictions.

the prediction error  $e_{ijt} \equiv g_{ijt} - g_{ijt}^e$  need not be zero, so variation in  $g_{ijt}$  reflects both variation in predictions and variation in prediction errors. Estimates that use  $g_{ijt}$  as the relevant regressor can therefore confound preference sensitivity  $\gamma$  with the unmodeled distribution of prediction errors. Section 6 and our empirical applications illustrate the resulting bias.

Dickstein and Morales (2018, Appendix A.4) establish that unknown information can prevent point identification in their export model. The following proposition demonstrates this failure in our model even when the full joint distribution of choices, observed attributes, and realized outcomes is known.

**Proposition 1** (Failure of point identification). *Under Assumptions 1–2, the preference vector  $\theta = (\beta', \gamma)'$  and the distribution of agents' predictions are not, in general, jointly point identified from the distribution of observed choices, attributes, and realized outcomes. This failure can arise even with homogeneous preferences, binary choice, and a known logit error distribution.*

*Proof.* See Appendix A.1. ■

The proof constructs a continuum of preference coefficients and informative signal structures that generate the same observable distribution. Rational expectations restrict prediction dispersion: with finite second moments,  $\text{Var}(g_{ijt}^e) \leq \text{Var}(g_{ijt})$ . They do not, however, uniquely determine predictions from realized outcomes. The next two sections impose additional structure to obtain point identification or informative bounds.

## 4 Point Identification with a Finite Mixture Model

This section develops a two-step finite-mixture approach. We use repeated realizations of the uncertain attribute to construct type-specific predictions from combinations of observed and recoverable latent predictors, then incorporate these predictions into a choice model to recover preferences and information-type shares.

### 4.1 Finite Mixture Model

To address the identification challenge posed by unobserved prediction information sets and information heterogeneity, we impose structure on the statistical process of the uncertain attribute  $g_{ijt}$ . This structure allows its observed and latent predictive components to be

estimated from repeated realizations of  $g_{ijt}$ , restricts the support of unobserved predictions  $g_{ijt}^e$ , and thereby disciplines the mapping from primitives to observed choice behavior.

Suppose the researcher observes a fixed list of candidate information variables  $K^{\text{obs}} = \{k_1, \dots, k_{|K^{\text{obs}}|}\}$  that are statistically related to the uncertain attribute  $g_{ijt}$ . For each product  $j$ , individual  $i$ , and market  $t$ , let

$$K_{ijt}^{\text{obs}} = (k_{1,ijt}, \dots, k_{|K^{\text{obs}}|,ijt}) \in \mathbb{R}^{|K^{\text{obs}}|}$$

denote the realized vector of these variables. We refer to the elements of  $K^{\text{obs}}$  as *information candidates*. Each candidate variable  $k \in K^{\text{obs}}$  satisfies the notion of *predictive information* in Definition 1(a):  $\mathbb{E}[g_{ijt} \mid k] \neq \mathbb{E}[g_{ijt}]$ .

The observed list  $K^{\text{obs}}$  need not contain every information variable used by agents. It provides the observed part of the statistical representation for  $g_{ijt}$ . Predictive variation not captured by this list may be summarized by fixed effects or latent factors that are unobserved by the researcher but recoverable from repeated realizations of  $g_{ijt}$ .

We allow individual prediction information sets  $\mathcal{W}_{it}$  to combine subsets of the observed candidates  $K_{ijt}^{\text{obs}}$  with these latent predictive components.

We specify the data-generating process (DGP) of the uncertain attribute as

$$g_{ijt} = g(K_{ijt}^{\text{obs}}; \tau) + v_{ijt}, \quad (3)$$

where  $g(\cdot; \tau)$  is an estimable function indexed by  $\tau$  and  $v_{ijt}$  contains latent predictive information and an unobserved shock. Equation (3) describes the statistical process of  $g_{ijt}$  and does not depend on agents' predictions or choices. For tractability, our exposition is based on a parameterized linear specification for  $g_{ijt}$ , but the first-stage function and the resulting conditional expectations may be estimated flexibly, as in our empirical applications.

We decompose the residual  $v_{ijt}$  to allow for latent predictors of  $g_{ijt}$  that agents may use when forming predictions. Two common specifications are an additive fixed-effect structure and a linear factor structure (as in CHN05):

$$v_{ijt} = v_i + v_j + v_t + \eta_{ijt} \quad (\text{additive fixed effects}), \quad (4)$$

$$v_{ijt} = \lambda_i' F_{jt} + \eta_{ijt} \quad (\text{linear factor model}). \quad (5)$$

In (5),  $\lambda_i \in \mathbb{R}^{|F|}$  denotes individual-specific loadings and  $F_{jt} \in \mathbb{R}^{|F|}$  denotes option and market factors, with  $|F|$  the number of factors. In either case,  $\eta_{ijt}$  is a residual satisfying

mean-independence given the corresponding predictors (e.g.,  $\mathbb{E}[\eta_{ijt} \mid K_{ijt}^{\text{obs}}, v_i, v_j, v_t] = 0$  in (4) and  $\mathbb{E}[\eta_{ijt} \mid K_{ijt}^{\text{obs}}, \lambda_i, F_{jt}] = 0$  in (5)).

For the remainder of the analysis, we work with an *augmented predictive basis* that includes both observed and latent predictors:

$$K_{ijt} \equiv (K_{ijt}^{\text{obs}}, K_{ijt}^{\text{unobs}}), \quad K_{ijt}^{\text{unobs}} = \begin{cases} (v_i, v_j, v_t), & \text{under (4),} \\ (\lambda_i, F_{jt}), & \text{under (5).} \end{cases}$$

Here “unobserved” means not directly observed by the researcher but recoverable from the data.<sup>4</sup> We focus on the linear factor model (5) in the rest of the paper; the arguments extend directly to the additive fixed-effect case.

Combining the specification choices above, we obtain the following data-generating process for the uncertain attribute:

$$g_{ijt} = K_{ijt}^{\text{obs}'} \tau + \lambda_i' F_{jt} + \eta_{ijt}. \quad (6)$$

Equation (6) parallels the role of outcome processes in CHN05 and the maintained model for  $g$  in DM18, which enable the researcher to characterize how expectations vary across alternatives.<sup>5</sup>

To represent information heterogeneity across individuals, we consider all feasible prediction information sets that can be formed from the augmented predictive basis. For each individual  $i$  and market  $t$ , let  $K_{it}$  denote the collection of observed and recovered predictive components obtained from  $\{K_{ijt}\}_{j \in \mathcal{J}}$ . Each element of  $K_{it}$  corresponds to one candidate predictor, which may itself vary across alternatives  $j$ .

Individual-specific prediction information sets  $\mathcal{W}_{it}$  are assumed to consist of subsets of these candidate variables. We therefore consider all non-empty subsets of the  $|K|$  candidate predictors as candidate prediction information sets used by different information types.<sup>6</sup>

Formally, let  $\mathcal{P}(K_{it})$  denote the collection of all non-empty subsets of the  $|K|$  candidate

---

<sup>4</sup>In a multifactor specification,  $\lambda_i$  and  $F_{jt}$  are treated as entire predictive blocks. Standard invertible factor rotations preserve conditional expectations based on those blocks. Treating individual factor coordinates as separate information candidates requires additional normalizations that identify their orientation.

<sup>5</sup>In practice, the functional form mapping predictors into  $g_{ijt}$  can be informed by the data using flexible statistical or machine learning methods, as in our empirical applications.

<sup>6</sup>The empty subset would generate an unconditional-expectation type with prediction  $g_{0,ijt}^e = \mathbb{E}[g_{ijt}]$ . It can be included in applications where its type-specific choice probabilities supply independent variation, subject to the rank conditions below.

variables, and index information types by  $\kappa \in \mathcal{K} \equiv \{1, \dots, |\mathcal{K}|\}$ , where

$$|\mathcal{K}| = |\mathcal{P}(K_{it})| = \sum_{k=1}^{|\mathcal{K}|} \binom{|\mathcal{K}|}{k} = 2^{|\mathcal{K}|} - 1 < \infty.$$

Each type  $\kappa$  is associated with a specific prediction information set  $\mathcal{W}_{\kappa,it} \in \mathcal{P}(K_{it})$  and a corresponding prediction

$$g_{\kappa,ijt}^e \equiv \mathbb{E}[g_{ijt} \mid \mathcal{W}_{\kappa,it}].$$

Let  $g_{\kappa,it}^e \equiv (g_{\kappa,ijt}^e)_{j \in \mathcal{J}}$  denote the vector of predictions generated by type  $\kappa$ , and let  $\mathcal{G}_{it} \equiv \{g_{\kappa,it}^e : \kappa \in \mathcal{K}\}$  denote the resulting finite prediction support. We keep the individual index  $i$  whenever information realizations, and hence predictions, vary across individuals within a type. If all individuals of type  $\kappa$  in market  $t$  have the same prediction for product  $j$ , then  $g_{\kappa,ijt}^e = g_{\kappa,jt}^e$  for every such individual, and we suppress  $i$  and write  $g_{\kappa,jt}^e$ . This framework nests a full-information benchmark in which one type conditions on the complete set of candidate predictors  $K_{it}$ , yielding the prediction  $\mathbb{E}[g_{ijt} \mid K_{it}]$  implied by the maintained data-generating process for  $g_{ijt}$ .

We now state separately the two restrictions that connect this statistical representation to the choice model.

**Assumption 3** (Recoverable finite prediction support).

- (i) **First-stage recoverability.** *The joint distribution of repeated observations of  $g_{ijt}$  and  $K_{ijt}^{\text{obs}}$  uniquely determines the finite collection of type-specific conditional-expectation functions whose evaluations form  $\mathcal{G}_{it}$ . If their construction uses latent predictive blocks, the blocks are identified up to transformations that preserve the conditional expectations.*
- (ii) **Finite mean-spanning.** *For every agent and market, the actual prediction vector  $g_{it}^e \equiv (g_{ijt}^e)_{j \in \mathcal{J}}$  satisfies  $g_{it}^e \in \mathcal{G}_{it}$  almost surely.*

Equation (6) as well as the additive fixed-effect specification provide a concrete first-stage model through which Assumption 3(i) can be satisfied when the repeated data and the applicable rank and normalization conditions identify the relevant latent blocks. Assumption 3(ii) imposes finite mean-spanning: for every agent and market, the vector of conditional means entering expected utility must belong to the finite support  $\mathcal{G}_{it}$  almost surely. Predictions outside this support are therefore excluded by assumption.

With the type-specific predictions  $g_{\kappa,ijt}^e$ , we model information heterogeneity using a finite-mixture specification. Let  $\kappa(i) \in \mathcal{K}$  denote the *information type* of individual  $i$ , drawn

from a type distribution

$$\phi = (\phi_1, \phi_2, \dots, \phi_{|\mathcal{K}|})' \in \mathbb{R}^{|\mathcal{K}|},$$

with probability mass function

$$\phi_\kappa \equiv \Pr(\kappa(i) = \kappa) \geq 0, \quad \kappa \in \mathcal{K}, \quad \sum_{\kappa=1}^{|\mathcal{K}|} \phi_\kappa = 1. \quad (7)$$

We maintain the following assumption.

**Assumption 4** (Finite-mixture of information types). *Every agent is of an information type in  $\mathcal{K}$ , with  $\kappa(i) \in \mathcal{K}$  drawn from the population distribution  $\phi$  defined in (7).*

Incorporating the finite-mixture structure into the demand system, both individual choice probabilities (with individual-level data) and market shares (with market-level data) can be written as mixtures of type-specific logit probabilities.<sup>7</sup> With individual-level data, the unconditional choice probability is

$$\Pr(d_{ijt} = 1) = \sum_{\kappa=1}^{|\mathcal{K}|} \phi_\kappa \frac{\exp(X'_{ijt}\beta + \gamma g_{\kappa,ijt}^e)}{1 + \sum_{j'=1}^J \exp(X'_{ij't}\beta + \gamma g_{\kappa,ij't}^e)}. \quad (8)$$

For the market-level specialization, suppose that  $X_{ijt} = X_{jt}$  and that type-specific predictions are common across individuals within a market, so that  $g_{\kappa,ijt}^e = g_{\kappa,jt}^e$ . The predicted share of product  $j$  in market  $t$  is then

$$s_{jt}(X_t, g_t^e; \Theta) = \sum_{\kappa=1}^{|\mathcal{K}|} \phi_\kappa s_{\kappa jt}(X_t, g_{\kappa,t}^e; \theta), \quad (9)$$

where

$$s_{\kappa jt}(X_t, g_{\kappa,t}^e; \theta) \equiv \frac{\exp(X'_{jt}\beta + \gamma g_{\kappa,jt}^e)}{1 + \sum_{j'=1}^J \exp(X'_{j't}\beta + \gamma g_{\kappa,j't}^e)}.$$

Here  $g_{\kappa,t}^e = (g_{\kappa,1t}^e, \dots, g_{\kappa,Jt}^e)'$  denotes the vector of predicted attributes for type  $\kappa$ , and  $\Theta \equiv (\theta', \phi')'$  collects preference parameters and mixture weights. If predictions instead vary across individuals within a type,  $s_{\kappa jt}$  is the cross-sectional average, conditional on

---

<sup>7</sup>Finite-mixture specifications have been used elsewhere to approximate the distribution of preference heterogeneity (e.g., Berry and Jia, 2010; Nevo et al., 2016; Bonhomme et al., 2022); here we instead use a finite mixture to approximate heterogeneity in predictions  $g_{ijt}^e$  induced by unobserved and heterogeneous information.

$\kappa(i) = \kappa$ , of the logit kernel in (8) using  $g_{\kappa,jt}^e$ . Figure A1 illustrates the construction of type-specific prediction information sets and their mapping into type-specific predictions.

An important benefit of the framework is that it allows us to test the content of agents' prediction information sets. We can test for the presence of a given type  $\kappa$  by assessing whether the corresponding mixture weight  $\phi_\kappa$  differs from zero.<sup>8</sup> A positive and significant estimate of  $\phi_\kappa$  implies that the information set  $\mathcal{W}_{\kappa,it}$  is used by a non-negligible share of the population when forming predictions.

The identified type-specific demand functions support counterfactuals that change prices, observed utility characteristics, or choice sets while maintaining the estimated information structure. They also support changes in the prevalence of information types. For a specified counterfactual distribution  $\tilde{\phi}$ , predicted shares are

$$\tilde{s}_{jt}(\hat{\theta}, \tilde{\phi}) = \sum_{\kappa \in \mathcal{K}} \tilde{\phi}_\kappa s_{\kappa jt}(\hat{\theta}). \quad (10)$$

Equation (10) maps a change in information use into choices. The researcher can specify  $\tilde{\phi}$  using experimental or institutional evidence on an information intervention or derive it from a model.

Our approach builds on CHN05. Their framework permits unobserved information factors and uses a common information set across individuals. We use the statistical process of  $g_{ijt}$  to recover observed and latent predictive components and allow the composition of prediction information sets to vary across individuals. DM18 builds prediction inequalities from a researcher-observed information set used by every agent. We construct an estimable support of conditional expectations from the augmented predictive basis. Given this first-stage structure, choice data recover the distribution of information types and the preference parameters.

Additionally, our framework allows for a discrete set of type-specific predictions  $\{g_{\kappa,jt}^e\}_{\kappa \in \mathcal{K}}$ . This accommodates heterogeneous information types and allows the data to discipline their prevalence by estimating the mixture weights  $\phi$  from observed choice patterns. We close this section with an example that illustrates our approach. The following example illustrates the construction of the information types and their corresponding predictions.

**Example 3 (Market-level Data).** Suppose the uncertain attribute is common across in-

---

<sup>8</sup>Because mixture weights are constrained to the simplex, the null  $\phi_\kappa = 0$  lies on the boundary of the parameter space. Conventional symmetric Wald inference need not be valid at this boundary, so estimates equal to zero are interpreted descriptively.

dividuals,  $g_{ijt} = g_{jt}$ , and admits the decomposition  $g_{jt} = k_{1,jt} + k_{2,jt} + \eta_{jt}$ , where  $k_{1,jt}$  is an observed predictor and  $k_{2,jt}$  is a latent factor recovered from repeated realizations of  $g_{jt}$ . The augmented predictive basis is therefore  $K_{ijt} = \{k_{1,jt}, k_{2,jt}\}$ , so  $|\mathcal{K}| = 2^2 - 1 = 3$  information types.

Let the three types be  $\kappa \in \{A, B, C\}$ , with type-specific information sets  $\mathcal{W}_{A,jt} = \{k_{1,jt}\}$ ,  $\mathcal{W}_{B,jt} = \{k_{2,jt}\}$ , and  $\mathcal{W}_{C,jt} = \{k_{1,jt}, k_{2,jt}\}$ . Their type-specific predictions are

$$g_{\kappa,jt}^e \equiv \mathbb{E}[g_{jt} \mid \mathcal{W}_{\kappa,jt}],$$

which imply

$$g_{A,jt}^e = k_{1,jt} + \mathbb{E}[k_{2,jt} \mid k_{1,jt}], \quad g_{B,jt}^e = k_{2,jt} + \mathbb{E}[k_{1,jt} \mid k_{2,jt}], \quad g_{C,jt}^e = k_{1,jt} + k_{2,jt}.$$

Type C corresponds to the full-information benchmark. The market share in (9) becomes

$$s_{jt}(X_t, g_t^e; \Theta) = \sum_{\kappa \in \{A, B, C\}} \phi_{\kappa} \frac{\exp(X'_{jt}\beta + \gamma g_{\kappa,jt}^e)}{1 + \sum_{j'=1}^J \exp(X'_{j't}\beta + \gamma g_{\kappa,j't}^e)}.$$

## 4.2 Identification

We focus on the more challenging case of market-level data  $\{s_t, X_t, g_t^e\}_{t \in \mathcal{T}}$ . Under Assumption 3(i), repeated realizations of  $g_t$  identify the finite support of type-specific predictions. Assumption 3(ii) links this first-stage support to choices by requiring it to span agents' actual conditional means. Let  $s_t = (s_{0t}, s_{1t}, \dots, s_{Jt})'$  denote observed shares in market  $t$ . Let  $s_{\kappa jt}(\theta)$  denote the type-specific choice probability implied by the structural parameter  $\theta$ . Stack these probabilities into the  $(J+1) \times |\mathcal{K}|$  matrix  $A_t(\theta)$  with entries  $(A_t(\theta))_{j\kappa} = s_{\kappa jt}(\theta)$ . The mixture structure implies the market share mapping

$$s_t = A_t(\theta)\phi.$$

For each candidate  $\theta$ , the first stage therefore gives us  $A_t(\theta)$ . The following rank conditions isolate the additional variation in choices required to identify preferences and type shares. We focus on local identification here to build intuition. Section 4.3 establishes global identification in a model with heterogeneous preferences under additional instrument, support, and rank conditions. This result also applies to the homogeneous model when these additional conditions are satisfied.

**Assumption 5.** The type-specific choice probability matrix  $A_t(\theta) \in \mathbb{R}^{(J+1) \times |\mathcal{K}|}$  has full column rank for at least one market  $t$ .

**Assumption 6.** The Jacobian matrix  $B_t(\theta_0, \phi_0) = \left( \frac{\partial s_t(\theta_0, \phi_0)}{\partial \theta}, A_t(\theta_0) \right) \in \mathbb{R}^{(J+1) \times (|\theta| + |\mathcal{K}|)}$  has full column rank for at least one market  $t$ .

**Theorem 1.** Consider the model defined by Assumptions 1–4. Under Assumption 5, the mixture weights  $\phi_0$  are uniquely identified given  $\theta_0$ .

If, in addition, Assumption 6 holds, then  $(\theta_0, \phi_0)$  is locally identified.

*Proof.* Assumption 3(i) identifies the type-specific predictions  $g_{\kappa, jt}^e$ . Assumption 3(ii), together with Assumption 4, implies that these predictions span the population choice probabilities. Hence, for any candidate  $\theta$ , the type-specific choice probabilities  $s_{\kappa jt}$  are identified functions and  $s_t(\theta, \phi) = A_t(\theta)\phi$ .

First, under Assumption 5,  $A_t(\theta_0)$  has full column rank. Hence the linear system  $s_t = A_t(\theta_0)\phi$  admits a unique solution for  $\phi$ , implying identification of  $\phi_0$  given  $\theta_0$ .

Second, define

$$F(\theta, \phi) \equiv s_t(\theta, \phi) - s_t = 0, \quad s_t(\theta, \phi) = A_t(\theta)\phi.$$

The Jacobian of  $F$  with respect to  $(\theta, \phi)$  at  $(\theta_0, \phi_0)$  is

$$B_t(\theta_0, \phi_0) = \left[ \frac{\partial s_t}{\partial \theta}(\theta_0, \phi_0), \frac{\partial s_t}{\partial \phi}(\theta_0, \phi_0) \right] = \left[ \frac{\partial s_t}{\partial \theta}(\theta_0, \phi_0), A_t(\theta_0) \right].$$

By Assumption 6, this matrix has full column rank. The implicit function theorem then implies that the solution to  $F(\theta, \phi) = 0$  is locally unique in a neighborhood of  $(\theta_0, \phi_0)$ , establishing local identification of  $(\theta_0, \phi_0)$ . ■

We present Assumptions 5 and 6 such that they rely solely on within-market variation. In particular, full column rank requires  $J + 1 \geq |\mathcal{K}|$  for  $A_t(\theta)$  and  $J + 1 \geq |\theta| + |\mathcal{K}|$  for  $B_t(\theta, \phi)$ . The rank condition for  $A_t(\theta)$  requires that type-specific choice probabilities  $s_{\kappa jt}(\theta)$  differ across  $\kappa$ , which in turn requires that the induced predictions  $g_{\kappa, jt}^e$  are not collinear and that  $\gamma \neq 0$ . If a single market does not provide sufficient variation, identification can instead rely on cross-market variation. Let  $\mathcal{A}(\theta)$  and  $\mathcal{B}(\theta, \phi)$  denote the matrices obtained by stacking  $A_t(\theta)$  and  $B_t(\theta, \phi)$  across markets. The rank conditions in Assumptions 5 and 6 then apply to these stacked matrices (which is generally easier to satisfy).

Theorem 1 characterizes local identification in a choice model with unobserved information heterogeneity and provides conditions for point identification. Because choices

depend on expected attributes, heterogeneity in information sets induces cross-sectional dispersion in predictions. Absent restrictions on the distribution of these predictions, observed choice variation cannot disentangle taste heterogeneity from belief heterogeneity, and preference parameters are not point identified.<sup>9</sup> We show that point identification can be restored by imposing structure on the data-generating process of the uncertain attribute, which restricts the support of prediction heterogeneity and generates estimable type-specific regressors. This yields a finite-mixture representation with an estimable support generated by observed and recovered latent predictors and permits explicit rank conditions for identification.

Our semiparametric finite-mixture approach relies on both clauses of Assumption 3: the finite prediction support must be recoverable from the statistical process of  $g$ , and it must span agents' actual conditional means. In Section 5, we relax these requirements: the researcher need not recover or enumerate the complete support of prediction rules. With weaker restrictions on the distribution of predictions, we obtain set identification of preference parameters rather than point identification. Finally, having established identification of  $(\theta, \phi)$  under the baseline finite-mixture specification, we now consider two extensions (i) market-specific information distributions and (ii) diagnostic tests of the information structure, before establishing identification in a more general model with endogenous product characteristics as well as preference heterogeneity.

**Extension 1** (Market-specific information types and shares). Identification of  $\phi$  conditional on  $\theta$  does not require cross-market variation. The mixture weights may therefore vary by market,  $\phi_t$ , provided that the rank condition on  $A_t(\theta)$  holds within each market. Likewise, the number of information types  $|\mathcal{K}|$  may differ across markets if the set of available information variables changes across markets. This extension permits the information distribution to vary with the market environment without altering the identification argument.

**Extension 2** (Testing for misspecification of information). Following CHN05, the maintained information representation can be formally tested under an explicit timing condition. Let  $\mathcal{F}_{it}^c$  contain all information available to individual  $i$  at the time of choice, with  $\mathcal{W}_{it} \subseteq \mathcal{F}_{it}^c$ , and suppose that the residual  $\eta_{ijt}$  in (6) is a nondegenerate ex-post innovation:

$$\mathbb{E}[\eta_{ijt} \mid \mathcal{F}_{it}^c] = 0, \quad \Pr(\text{Var}(\eta_{ijt} \mid \mathcal{F}_{it}^c) > 0) > 0.$$

---

<sup>9</sup>See Barseghyan et al. (2021); Molinari (2020); Agarwal and Somaini (2026) for related approaches that separate preferences and beliefs in other environments; unlike our framework, these papers focus on settings with unobserved choice sets rather than unobserved information over attributes.

This timing condition implies that no individual observes  $\eta_{ijt}$ , or any signal that predicts it in conditional mean, when choosing. In particular, it rules out perfect foresight.

After estimating the DGP for  $g_{ijt}$ , the researcher can add a diagnostic type  $\kappa_\eta$  whose information set augments  $K_{it}$  with  $\{\hat{\eta}_{ijt}\}_{j \in \mathcal{J}}$ , giving the prediction  $\hat{g}_{\kappa_\eta,ijt}^e = g_{ijt}$ . Under the maintained DGP and timing condition, this type is infeasible, so the model implies

$$H_0 : \phi_{\kappa_\eta} = 0.$$

A test of this null, using inference appropriate for a boundary restriction, is therefore a formal specification test. Rejection indicates that the residual contains choice-relevant variation not spanned by the maintained information representation. This mirrors the logic in CHN05: once all choice-relevant information is included, remaining variation in the uncertain attribute should be orthogonal to choices.<sup>10</sup>

### 4.3 Preference heterogeneity, endogeneity, and choice-specific unobservables

We extend the framework to allow for unobserved preference heterogeneity, endogenous characteristics, and choice-specific unobservables. Following Berry and Haile (2014), we work with market-level data. Let  $x_{jt}^{(1)}$  denote exogenous characteristics,  $p_{jt}$  an endogenous attribute (e.g., price),  $\xi_{jt}$  an unobserved demand shock, and  $z_{jt}$  excluded instruments satisfying the mean-independence restriction below. Let  $\tilde{x}_{jt}^{(2)}$  denote characteristics with heterogeneous tastes. We also allow for individual-specific sensitivity  $\gamma_i$  to the predicted attribute  $g_{\kappa(i),jt}^e$ .

Expected utility becomes

$$u_{ijt}^e = x_{jt}^{(1)'} \beta - \alpha p_{jt} + \gamma_i g_{\kappa(i),jt}^e + \tilde{x}_{jt}^{(2)'} \beta_i^{(2)} + \xi_{jt} + \epsilon_{ijt}. \quad (11)$$

The random coefficients  $(\beta_i^{(2)}, \gamma_i)$  follow a distribution  $\tilde{F}_{\beta^{(2)}, \gamma}$ .<sup>11</sup> For identification, write  $\tilde{\delta}_{jt} = x_{jt}^{(1)'} \beta + \xi_{jt}$  and  $\delta_{jt} = \tilde{\delta}_{jt} - \alpha p_{jt}$ . The Berry–Haile representation keeps prices separate from  $\tilde{\delta}_t$ , while the mixed-logit representation uses  $\delta_t$ .

<sup>10</sup>Without the timing condition, some agents may observe  $g_{ijt}$  or a signal correlated with the residual, so a positive weight may represent genuinely richer information rather than misspecification. If the ex-post innovation is absent, the residual adds no variation and the diagnostic is uninformative.

<sup>11</sup>The framework can accommodate a random price coefficient by including  $\alpha_i$  in the random-coefficient vector, subject to corresponding support and rank conditions. The identification argument below maintains a homogeneous price coefficient.

Let  $\tilde{g}_{jt}^e \equiv (g_{1,jt}^e, \dots, g_{|\mathcal{K}|,jt}^e)' \in \mathbb{R}^{|\mathcal{K}|}$  collect the type-specific predictions for option  $j$  in market  $t$ . The information structure restricts the prediction component to a finite set of known directions. This transforms Assumption 1 into:

**Assumption 1'** (Choice under uncertainty'). *At the time of choice, individuals are uncertain only about  $g_{ijt}$  and form rational expectations conditional on  $\mathcal{W}_{it}$ . They choose the option that maximizes expected utility given in (11). Taste shocks  $(\epsilon_{i0t}, \dots, \epsilon_{ijt})$  are i.i.d. Type-I extreme value with scale normalized to one and are independent of market characteristics, demand shocks, and information sets.*

To apply Berry and Haile (2014) and Wang (2023), we express the prediction component as a random coefficient on the vector of type-specific predictions:

**Proposition 2.** *Under Assumption 1' and Assumptions 2–4, the model (11) can be written as a standard random utility model with utility*

$$u_{ijt}^e = x_{jt}^{(1)'} \beta - \alpha p_{jt} + x_{jt}^{(2)'} \beta_i + \xi_{jt} + \epsilon_{ijt},$$

where  $x_{jt}^{(2)}$  stacks  $\tilde{x}_{jt}^{(2)}$  and  $\tilde{g}_{jt}^e$ , and  $\beta_i = (\beta_i^{(2)'}, \tilde{\gamma}_i')'$  has distribution  $F \equiv F_{\beta^{(2)}, \tilde{\gamma}}$ .

The coefficient on the prediction component admits the representation

$$\tilde{\gamma}_i = \gamma_i b_{\kappa(i)},$$

where  $\gamma_i$  is an individual-specific scale parameter and  $\kappa(i) \in \mathcal{K}$  denotes the information type drawn from the distribution  $\phi$ . Each  $b_\kappa$  is the  $\kappa$ th canonical basis vector, so these known vectors determine the finite set of directions induced by the information structure.

*Proof.* See Appendix A.2. ■

**Assumption 7** (Exogeneity of random coefficients and information types). *The joint vector  $(\beta_i^{(2)}, \gamma_i, \kappa(i))$  is independent of  $(x_t^{(1)}, x_t^{(2)}, p_t, \xi_t, \epsilon_{it})$ , where  $x_t^{(2)}$  includes all type-specific prediction regressors and  $\epsilon_{it} = (\epsilon_{i0t}, \dots, \epsilon_{ijt})$ . Moreover,  $(\beta_i^{(2)}, \gamma_i) \perp \kappa(i)$ , with distributions  $\tilde{F}_{\beta^{(2)}, \gamma}$  and  $\phi$ , respectively, and  $\Pr(\gamma_i \neq 0) > 0$ .*

Assumption 7 makes the induced coefficient distribution independent of market characteristics, including all prediction regressors, as in Wang (2023). The common-tastes condition  $(\beta_i^{(2)}, \gamma_i) \perp \kappa(i)$  rules out sorting by preferences and allows type shares  $\phi$  to be recovered from relative masses on the coordinate axes.<sup>12</sup>

<sup>12</sup>Similar independence assumptions appear in the literature on limited consideration where relaxing the

If preference distributions vary across types, the relevant primitives become  $\{F_{\beta^{(2)}, \gamma | \kappa}, \phi_\kappa\}_{\kappa \in \mathcal{K}}$ , and the factorization used by the current identification result and estimator no longer applies.<sup>13</sup> Because the direction  $b_{\kappa(i)}$  reveals the information type when  $\gamma_i \neq 0$ , independence is sufficient rather than necessary for this separation. Establishing identification conditions and an estimator for type-specific preference distributions is left for future work.<sup>14</sup>

**Theorem 2.** *Under Assumptions 1', 2–4, and 7, the Berry–Haile conditions BH1–BH4, the Wang conditions W1–W4, and the IV rank condition in Assumption 10, all stated in Appendix A.3, the primitives  $(\beta_0, \alpha_0, \phi_0, \tilde{F}_{\beta^{(2)}, \gamma})$  in model (11) are globally identified.*

*Proof.* See Appendix A.3. ■

The argument separates identification of demand from identification of the coefficient distribution. The conditions of Berry and Haile (2014) identify demand on its support. Instrument exogeneity requires  $\mathbb{E}[\zeta_{jt} \mid z_t, x_t^{(1)}, x_t^{(2)}] = 0$ , where  $x_t^{(2)}$  includes all type-specific predictions.

Wang’s result requires demand to be identified on an open set of  $\delta_t$  at a full-column-rank matrix of random-coefficient and prediction regressors. Under these conditions, his theorem globally identifies the induced coefficient distribution. We then recover our model’s structural primitives: a linear-IV rank condition identifies the homogeneous coefficients, while the known coordinate structure and common-tastes restriction recover the taste distribution and information-type shares. The appendix states the conditions for this final step explicitly.

## 4.4 Estimation

Both estimation procedures use first-stage estimates of the candidate prediction functions. We require these estimates to be uniformly consistent over the finite collection of candidate

---

independence between attention and utility shocks generally requires additional instruments (e.g., Abaluck and Adams-Prassl, 2021).

<sup>13</sup>If  $\gamma_i$  can equal zero with positive probability, the coefficient direction does not reveal the information type for these individuals. The identified off-origin mass satisfies  $m_\kappa \equiv \Pr(\tilde{\gamma}_i \in \{cb_\kappa : c \in \mathbb{R} \setminus \{0\}\}) = \phi_\kappa \Pr(\gamma_i \neq 0 \mid \kappa)$ . Without restrictions on the type-specific probability of  $\gamma_i = 0$ ,  $m_\kappa$  does not separately identify  $\phi_\kappa$ .

<sup>14</sup>Barseghyan et al. (2021) explicitly allow unrestricted dependence (conditional on observables) between unobserved choice sets and preferences, and obtain partial identification via moment inequalities. Brown and Jeon (2024) also allow for endogenous sorting into information types with a full structural model of rational inattention.

functions; a high-level sufficient condition is

$$\max_{\kappa \in \mathcal{K}} \|\hat{g}_\kappa^e - g_\kappa^e\|_{L_2} = o_p(1).$$

Because these predictions are generated in a first stage, inference must account for their sampling uncertainty.

**Estimation with Individual-level Data.** Given individual-level data, we estimate the parameters of interest by maximum likelihood. Let  $d_{ijt}$  denote an indicator equal to one if individual  $i$  chooses product  $j$  in market  $t$ , with  $\sum_{j=0}^J d_{ijt} = 1$ . The likelihood function is

$$\mathcal{L}(\Theta|d) = \prod_{i,t} \prod_{j=0}^J [\Pr(d_{ijt} = 1 \mid \Theta)]^{d_{ijt}}. \quad (12)$$

The choice probabilities  $\Pr(d_{ijt} = 1 \mid \Theta)$  correspond to the mixture logit probabilities defined in Equation (8). When the choice set contains only a single product ( $J = 1$ ), the multinomial likelihood reduces to the binary-choice likelihood used in DM18.

**Estimation with Market-level Data.** With market-level data, we extend the GMM approach of Berry et al. (1995). Theorem 2 establishes identification of the augmented random-coefficient demand system. In the empirical specifications, the random-coefficient distribution is parameterized by a finite-dimensional vector  $\Theta$ . For each candidate  $\Theta$ , the type-specific predictions enter as additional random-coefficient regressors, and the standard BLP share inversion recovers the mean utilities and structural errors  $\xi_{jt}(\Theta)$ .

Let  $\xi_{jt}(\Theta)$  denote the structural error (mean utility residual) implied by the model at parameter vector  $\Theta = (\theta', \phi')'$ . We impose mean independence of  $\xi_{jt}(\Theta)$  with respect to a vector of instruments  $Z_{jt} \in \mathbb{R}^{|Z|}$ :

$$\mathbb{E}[\xi_{jt}(\Theta_0) \mid Z_{jt}] = 0,$$

where  $|Z|$  denotes the number of instruments. The instrument vector  $Z_{jt}$  stacks (i) information shifters used to construct type-specific predictions, (ii) exogenous observed attributes, and (iii) additional excluded instruments for endogenous components of  $X_{jt}$  (e.g., cost shifters for prices).

The estimator chooses  $\Theta$  to minimize a quadratic form in the stacked sample moments. Let  $\xi(\Theta)$  and  $Z$  denote the vectors/matrices stacked over  $(j, t)$ , and let  $W$  be a (consistent)

estimate of the optimal weighting matrix. We solve

$$\begin{aligned} \min_{\Theta} \quad & \left( \frac{1}{N} Z' \xi(\Theta) \right)' W \left( \frac{1}{N} Z' \xi(\Theta) \right) \\ \text{s.t.} \quad & \phi_{\kappa} \geq 0 \quad \forall \kappa \in \mathcal{K}, \quad \sum_{\kappa \in \mathcal{K}} \phi_{\kappa} = 1, \end{aligned}$$

where  $N$  is the number of  $(j, t)$  observations used in the moments.

In principle, the information shifters  $\{k_{mjt}\}$  are potential instruments. In practice, however, they often perform poorly because they enter the moments only through the nonlinear mapping from information to predicted attributes and then to shares. Moreover, the researcher may include candidate information variables (or types) that are irrelevant because they are not used by any agents, further weakening standard IV variation.

To address both issues, we use Chamberlain's approximation to optimal instruments (Chamberlain, 1987), which is proportional to the conditional expectation of the score/Jacobian,  $\mathbb{E}[\partial \xi_{jt}(\Theta) / \partial \Theta' \mid Z_{jt}]$ . These instruments (i) apply a nonlinear operator to the baseline instrument set  $Z_{jt}$ , capturing the nonlinear way information affects shares, and (ii) target the parameters of interest directly (notably  $\phi$  and  $\gamma$ ). If a candidate type  $\kappa$  is irrelevant, then  $\partial \xi_{jt}(\Theta) / \partial \phi_{\kappa}$  is approximately zero, so the corresponding optimal instrument contributes negligibly, effectively down-weighting nonexistent types.

Following Reynaert and Verboven (2014), we implement a two-stage approximation. In the first stage we estimate the model using baseline instruments  $Z_{jt}$  (including information shifters and standard demand instruments). We then construct empirical Chamberlain instruments by evaluating  $\mathbb{E}[\partial \xi_{jt}(\Theta) / \partial (\phi', \gamma)' \mid Z_{jt}]$  at the first-stage estimates. In the second stage, we either (i) run a standard two-step GMM that treats the constructed Chamberlain instruments as fixed and uses an updated optimal weighting matrix, or (ii) use a continuously updated GMM (CUE) estimator that recomputes the Chamberlain instruments over the minimization. In our applications, the CUE approach is often more stable in mixture models because it keeps the instruments aligned with the current parameter values along the optimization path.<sup>15</sup>

---

<sup>15</sup>When the number of candidate information types is large, the linear-mixture estimator of Fox et al. (2011) and its elastic-net extension in Heiss et al. (2022) offer computationally convenient alternatives.

## 5 Moment Inequalities from Extremal Information

We now develop a partial-identification approach for inference on preference parameters that relaxes the recoverability and enumeration requirements in Assumption 3 and is feasible in the presence of many types. The prediction distribution remains finite, but the researcher need not identify its complete support. The key idea is simple: even if we do not observe agents' predictions  $g_{jt}^e$ , revealed preference implies bounds on preferences once we can isolate states in which an *observed* prediction rule delivers extremal prediction errors. In those extreme states, the observed rule provides a worst-case (or best-case) approximation to the unobserved prediction entering utility, generating moment inequalities that are valid under heterogeneous and unobserved prediction information sets. This approach is implementable with market-level data, whereas existing moment-inequality methods such as DM18 are developed for individual-level data.

Formally, we show that partial knowledge of the cross-sectional distribution of prediction errors—in particular, its behavior in the tails conditional on observable information shifters—is sufficient to set identify  $\theta$ . Operationally, the researcher needs (i) to observe the prediction information set of at least one information type that is present in the data, and (ii) to construct a selection rule that picks observations in which this observed type attains the maximal or minimal prediction error among all types. These extremal events allow us to replace unknown bounds driven by unobserved information with observable bounds driven by the selected type, yielding inequalities that are informative about  $\theta$ .

### 5.1 Identification

We start by defining conditional moment inequalities given a vector of conditioning variables and instruments. Let  $Z_{jt}$  denote the vector of variables observed by the researcher and used for conditioning. It may include (i) shifters used to form observable predictions  $g_{\tilde{\kappa}jt}^e$  for an observed information type  $\tilde{\kappa}$ , and (ii) standard demand instruments used to address endogeneity in  $X_{jt}$  through the unobserved demand shock  $\zeta_{jt}$ . We consider the conditional inequalities

$$\mathbf{m}_{\max}(Z_{jt}; \theta) \geq 0, \quad \mathbf{m}_{\min}(Z_{jt}; \theta) \geq 0, \quad (13)$$

where

$$\begin{aligned}\mathbf{m}_{\max}(Z_{jt}; \theta) &\equiv \mathbb{E} \left[ -\log \left( \frac{s_{jt}}{s_{0t}} \right) + X'_{jt}\beta + \gamma g_{jt} + \zeta_{jt} + \max_{\kappa \in \mathcal{K}_t} \{-\gamma e_{\kappa jt}\} \mid Z_{jt} \right], \\ \mathbf{m}_{\min}(Z_{jt}; \theta) &\equiv \mathbb{E} \left[ \log \left( \frac{s_{jt}}{s_{0t}} \right) - X'_{jt}\beta - \gamma g_{jt} - \zeta_{jt} - \min_{\kappa \in \mathcal{K}_t} \{-\gamma e_{\kappa jt}\} \mid Z_{jt} \right].\end{aligned}\tag{14}$$

The population inequalities above are valid for either sign of  $\gamma$ , but the sign determines which extremal-error event implements each bound. To write a single pair of implementable inequalities without repeatedly relabeling these events, we maintain the following restriction.

**Assumption 8** (Known sign of the expected-attribute coefficient). *The sign of  $\gamma$  is known a priori. In the maintained cost/expense normalization, the parameter space restricts  $\gamma \in \Gamma \subset (-\infty, 0)$ . If the uncertain attribute is instead a nonnegative benefit or revenue, the analogous restriction is  $\gamma \in \Gamma \subset (0, \infty)$ , and the maximal- and minimal-error events below are interchanged.*

Assumption 8 is not required for the population bounds or for set identification. Without it, the researcher can partition the candidate space into  $\Gamma_- \equiv \Gamma \cap (-\infty, 0)$  and  $\Gamma_+ \equiv \Gamma \cap (0, \infty)$ , use the appropriate extremal-event mapping within each region, treat  $\gamma = 0$  directly, and take the union of the resulting identified sets. We impose the restriction only to simplify notation and implementation. It is also natural in our applications, where the uncertain attribute has a known economic role as either a cost/expense or a benefit/revenue.

Let  $\Theta$  denote the parameter space for  $\theta = (\beta', \gamma)'$ . Define the identified set

$$\Theta_0 \equiv \left\{ \theta \in \Theta : \mathbf{m}_{\max}(Z_{jt}; \theta) \geq 0 \text{ and } \mathbf{m}_{\min}(Z_{jt}; \theta) \geq 0 \text{ for almost every } Z_{jt} \right\}.$$

**Theorem 3.** *Assume a finite-mixture logit structure. Then the population conditional inequalities in (14) hold. In particular, the true parameter  $\theta_0 \in \Theta_0$ .*

Theorem 3 states that the inequalities are satisfied at the true parameter value. The proof is in Appendix A.4. The intuition is clearest from the equivalent bounds on the observed log share ratio  $\Delta_{jt} \equiv \log(s_{jt}/s_{0t})$ :

$$\gamma g_{jt} + X'_{jt}\beta + \zeta_{jt} + \min_{\kappa \in \mathcal{K}_t} \{-\gamma e_{\kappa jt}\} \leq \Delta_{jt} \leq \gamma g_{jt} + X'_{jt}\beta + \zeta_{jt} + \max_{\kappa \in \mathcal{K}_t} \{-\gamma e_{\kappa jt}\}.\tag{15}$$

In words, for fixed  $(X_{jt}, g_{jt})$ , the log-odds must lie between the smallest and largest utility

indices induced by feasible prediction errors across information types.<sup>16</sup>

Under the maintained cost normalization  $\gamma < 0$ , the lower and upper indices in (15) can be written as

$$L_{jt}(\gamma) = X'_{jt}\beta + \xi_{jt} + \gamma \max_{\kappa \in \mathcal{K}_t} g_{\kappa,jt}^e, \quad U_{jt}(\gamma) = X'_{jt}\beta + \xi_{jt} + \gamma \min_{\kappa \in \mathcal{K}_t} g_{\kappa,jt}^e.$$

The sign restriction determines the relevant extrema, but it does not by itself sign the derivatives of these bounds. If the predicted attribute is a nonnegative cost or expense, then  $\partial L_{jt}(\gamma)/\partial \gamma = \max_{\kappa} g_{\kappa,jt}^e \geq 0$  and  $\partial U_{jt}(\gamma)/\partial \gamma = \min_{\kappa} g_{\kappa,jt}^e \geq 0$ . In that case, the nonnegative right-bound residual  $U_{jt}(\gamma) - \Delta_{jt}$  delivers a lower bound on  $\gamma$ , whereas the left-bound residual  $\Delta_{jt} - L_{jt}(\gamma)$  delivers an upper bound. If predictions can take either sign, we make no monotonicity claim and evaluate both inequalities directly at each candidate  $\gamma \in \Gamma$ .

Inequalities (15) are not directly implementable because they involve the extremal prediction errors  $\max_{\kappa} e_{\kappa jt}$  and  $\min_{\kappa} e_{\kappa jt}$ . We therefore work with *extremal-error events* that allow the researcher to replace these extrema by the prediction error of an observed information type.<sup>17</sup>

**Definition 2** (Extremal-error indices). Let  $e_{\kappa jt} \equiv g_{jt} - g_{\kappa jt}^e$  denote the prediction error of type  $\kappa$  for option  $j$  in market  $t$ . Define

$$\kappa_{\max}(j, t) \in \arg \max_{\kappa \in \mathcal{K}_t} e_{\kappa jt}, \quad \kappa_{\min}(j, t) \in \arg \min_{\kappa \in \mathcal{K}_t} e_{\kappa jt}.$$

These indices locate the types that attain the maximal/minimal prediction error.

Rather than attempting to recover  $\kappa_{\max}(j, t)$  and  $\kappa_{\min}(j, t)$  for each observation, we assume the researcher observes the prediction information set for at least one information type  $\tilde{\kappa}$  that exists in the data, and can construct selection indicators that isolate observations where the observed  $\tilde{\kappa}$  attains the maximal or minimal error.

Let  $Z_{jt}^P$  denote standard demand instruments for endogenous components of  $X_{jt}$  (e.g., cost shifters). We collect conditioning variables in

$$Z_{jt} \equiv (Z_{jt}^P, X_{jt}, \mathcal{W}_{\tilde{\kappa} jt}).$$

<sup>16</sup>The inequalities are closely related to the linear conditional moment-inequality approach in Gandhi et al. (2023), who use market-level data to conduct robust inference in settings with zero market shares.

<sup>17</sup>Extremality refers to the realized prediction error  $e_{\kappa jt}$  and does not imply that the corresponding type has more or less accurate information in general.

**Assumption 9** (Extremal-error selection). *Assume:*

- (i) *The researcher observes  $\mathcal{W}_{\tilde{\kappa}jt}$  for at least one information type  $\tilde{\kappa}$  that exists in the data.*
- (ii) *There exists an indicator  $Z_{\max jt}^S \in \{0, 1\}$  such that, on selected observations,  $\tilde{\kappa}$  attains the maximal prediction error in the sense that*

$$\mathbb{E} \left[ e_{\tilde{\kappa}jt} Z_{\max jt}^S \mid Z_{jt} \right] = \mathbb{E} \left[ \max_{\kappa \in \mathcal{K}_t} e_{\kappa jt} Z_{\max jt}^S \mid Z_{jt} \right].$$

- (iii) *There exists an indicator  $Z_{\min jt}^S \in \{0, 1\}$  such that, on selected observations,  $\tilde{\kappa}$  attains the minimal prediction error in the sense that*

$$\mathbb{E} \left[ e_{\tilde{\kappa}jt} Z_{\min jt}^S \mid Z_{jt} \right] = \mathbb{E} \left[ \min_{\kappa \in \mathcal{K}_t} e_{\kappa jt} Z_{\min jt}^S \mid Z_{jt} \right].$$

- (iv) *The unobserved demand shock has conditional mean zero on the selected observations:*

$$\mathbb{E} \left[ \tilde{\xi}_{jt} Z_{\max jt}^S \mid Z_{jt} \right] = 0, \quad \mathbb{E} \left[ \xi_{jt} Z_{\min jt}^S \mid Z_{jt} \right] = 0.$$

Assumption 9 is the key identifying restriction that turns the extremal bounds in (14) into feasible moment inequalities. Its validity is application-specific: the observed type must attain the relevant extremum on the selected observations, and selection must preserve the conditional mean independence of the demand shock. Because the predictions of the other types are unobserved, extremality cannot generally be verified directly in applications. The threshold rule below provides an operational way to construct the selection events, while Section 6 illustrates how threshold choices and classification errors affect the resulting confidence sets.

A natural selection rule uses thresholds on the observed-type error:

$$Z_{\max jt}^S = \mathbb{1}\{e_{\tilde{\kappa}jt} \geq a_{\max}\}, \quad Z_{\min jt}^S = \mathbb{1}\{e_{\tilde{\kappa}jt} \leq a_{\min}\},$$

where the thresholds can be chosen as extreme quantiles of the empirical distribution of  $e_{\tilde{\kappa}jt}$ . This choice entails a validity–power trade-off. Tighter thresholds aim to reduce non-extremal selections, supporting the validity of the moment inequalities, but use fewer observations and may reduce statistical power. Extreme thresholds alone do not guarantee Assumption 9.

Applying these selection indicators to (14) yields implementable conditional moment

inequalities that depend only on the observed prediction  $g_{\tilde{\kappa}jt}^e$ :

$$\begin{aligned} \mathbf{m}_{\max}^{\tilde{\kappa}}(Z_{jt}; \theta) &\equiv \mathbb{E} \left[ \left( -\log \frac{s_{jt}}{s_{0t}} + \gamma g_{\tilde{\kappa}jt}^e + X'_{jt}\beta + \zeta_{jt} \right) \cdot Z_{\max jt}^S \mid Z_{jt} \right] \geq 0, \\ \mathbf{m}_{\min}^{\tilde{\kappa}}(Z_{jt}; \theta) &\equiv \mathbb{E} \left[ \left( \log \frac{s_{jt}}{s_{0t}} - \gamma g_{\tilde{\kappa}jt}^e - X'_{jt}\beta - \zeta_{jt} \right) \cdot Z_{\min jt}^S \mid Z_{jt} \right] \geq 0, \end{aligned} \quad (16)$$

where  $g_{\tilde{\kappa}jt}^e \equiv \mathbb{E}[g_{jt} \mid \mathcal{W}_{\tilde{\kappa}jt}]$  is the prediction implied by the observed information type  $\tilde{\kappa}$ . Under the maintained cost normalization  $\gamma < 0$ , the maximal-error event supplies the upper utility-index bound and the minimal-error event supplies the lower bound, as written in (16). For a benefit or revenue with  $\gamma > 0$ , these two error events are interchanged.

**Example 2 (continued)** (Automobile Purchase). Consider the fuel-cost setting of Example 2, where the uncertain attribute is realized fuel cost per mile  $g_{jt} = e_{jt}f_t$  and consumers form predictions  $g_{jt}^e = \mathbb{E}[g_{jt} \mid \mathcal{W}]$  from subsets of the observable predictors  $\{e_{jt}, f_t\}$ . The generic coefficient on the predicted cost is  $-\alpha_i\gamma\rho < 0$ , consistent with Assumption 8. Suppose there exists an information type  $\tilde{\kappa}$  that predicts fuel costs using fuel consumption only,  $\mathcal{W}_{\tilde{\kappa}} = \{e_{jt}\}$ . When fuel prices are sufficiently high, this type's omitted fuel-price information generates large prediction errors. The extremal-selection condition required for (16) is that, on a subset of such high-price observations, type  $\tilde{\kappa}$  attains the maximal prediction error among all types.

## 5.2 Estimation

The estimation of the moment-inequality approach restricts the candidate parameter space to the economically known sign in Assumption 8. We evaluate both moments directly for every candidate  $\gamma \in \Gamma$ ; monotonicity is used only when nonnegative predictions justify the interval interpretation above. The estimation then proceeds in three steps. First, we fix an observed information type indexed by  $\tilde{\kappa} \in \mathcal{K}$  and specify its information set  $\mathcal{W}_{\tilde{\kappa}jt}$ . Using this information set, we estimate the type-specific predictions  $g_{\tilde{\kappa}jt}^e = \mathbb{E}[g_{jt} \mid \mathcal{W}_{\tilde{\kappa}jt}]$  and the associated prediction errors  $e_{\tilde{\kappa}jt} = g_{jt} - g_{\tilde{\kappa}jt}^e$ . Second, we construct a selection indicator  $Z_{jt}^S$  that targets observations with extreme values of  $e_{\tilde{\kappa}jt}$ . The choice of  $Z_{jt}^S$  entails a validity-power trade-off. More extreme cutoffs can make the extremal-selection restriction more plausible but reduce the effective sample size, potentially weakening power to reject false parameter values and widening confidence sets. Looser cutoffs retain more observations and can increase power when the moments remain valid. However, non-extremal selections may invalidate the moment inequalities and compromise confidence-set coverage. Section 6 illustrates the sensitivity of the confidence sets to this choice. Third, we

adapt the conditional moment-inequality inference framework of Andrews and Shi (2013) to construct confidence regions for the preference parameters based on the inequalities in (16).<sup>18</sup>

## 6 Simulations

We use simulations with market-level data to assess the performance of the finite-mixture estimator and the sensitivity of moment-inequality confidence sets to the selection rule.

### 6.1 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  $T = 150$  markets and  $J = 10$  products. The indirect utility is defined as

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

The product characteristic is uniformly distributed  $x_{jt} \stackrel{\text{iid}}{\sim} U(1, 2)$ . We simulate the price  $p_{jt} = 1 + z_{jt} + v_{jt}$  to be exogenous with a cost shifter  $z_{jt} \stackrel{\text{iid}}{\sim} U(0, 1)$  and a cost shock  $v_{jt} \stackrel{\text{iid}}{\sim} U(-0.25, 0.25)$ . The uncertain attribute is assumed to be constant across individuals  $g_{ijt} = g_{jt}$  and equals the sum of two information variables and one random error  $g_{jt} = k_{1jt} + k_{2jt} + \varepsilon_{jt}^g$ . The three components are independent of each other. The first information variable is assumed to be observed by the researcher and follows a log-normal distribution  $\log(k_{1jt}) \stackrel{\text{iid}}{\sim} N(0, 1)$ . The second information variable is unobserved by the researcher and follows a linear factor model  $k_{2jt} = \lambda_j' F_t$  with  $\lambda_j \stackrel{\text{iid}}{\sim} N(1, 1)$  and an AR(1) process  $F_t = 0.5F_{t-1} + \varepsilon_t^F$ ,  $\varepsilon_t^F \stackrel{\text{iid}}{\sim} N(0, 1)$ . The random shock to the uncertain attribute follows a normal distribution  $\varepsilon_{jt}^g \stackrel{\text{iid}}{\sim} N(0, 0.2^2)$ . The demand shifters are uniformly distributed  $\xi_{jt} \stackrel{\text{iid}}{\sim} U(-1, 1)$ . The idiosyncratic demand shocks  $\epsilon_{ijt}$  follow a standard 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 proportions of the information types are  $\phi_A = 0.15$ ,  $\phi_B = 0.5$ , and  $\phi_C = 0.35$ . Here, type C

---

<sup>18</sup> Andrews et al. (2023) and Cox and Shi (2023) provide alternative inference procedures for linear conditional moment inequalities; we use Andrews and Shi (2013) throughout.

is a full-information type, and type  $B$  uses only the information variable that is unobserved by the researcher.

We also provide simulations where we introduce preference heterogeneity through random coefficients on the valuation of expected attributes  $\gamma$ . The random coefficient is specified as  $\gamma_i = \gamma + \sigma_\gamma v_i$ , where  $\gamma = -1.5$  and  $\sigma_\gamma = 0.5$ . We approximate this distribution using 100 Halton draws transformed to standard-normal variates, so that  $\gamma_i \sim N(-1.5, 0.5^2)$ .

Finally, we conduct a third simulation using datasets that contain only two information types,  $\kappa \in \{B, C\}$ , with shares  $\phi_B = 0.4$  and  $\phi_C = 0.6$ , to test whether our method can recover consistent estimates when there are fewer information types in the data than specified. We test the method in the most challenging case, where the DGP does not contain type  $A$ , which uses only the observed information variable. 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 to  $x_{jt} \stackrel{\text{iid}}{\sim} U(0, 1)$ .

## 6.2 Simulation of the Finite-Mixture Approach

We estimate the model without preference heterogeneity using the two-step GMM procedure described in Section 4.4. Specifically, we employ as regular instruments constant,  $x_{jt}$ ,  $p_{jt}$ ,  $z_{jt}$ ,  $k_{1jt}$  and (estimated)  $k_{2jt}$  in the first stage. This preliminary instrument set is not by itself sufficient to separately identify all parameters in the model, but the role of the first stage is to deliver initial estimates used to construct feasible optimal instruments.<sup>19</sup> Following Reynaert and Verboven (2014), we calculate approximate Chamberlain instruments for  $\gamma$  and  $\phi$  based on the first-stage estimates. We then replace the information variables  $k_{1jt}$ ,  $k_{2jt}$  in the set of regular instruments with Chamberlain instruments in the second stage.

Table 1 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 instruments is smaller than that of estimates using regular instruments. Especially for the intercept, information parameters  $\phi$ , and valuation of the expected attribute  $\gamma$ , the bias is significantly reduced when using Chamberlain instruments. In columns (9)–(11), we precisely estimate the information-type shares and

---

<sup>19</sup>When estimating the model on datasets with preference heterogeneity that contain only information types  $B$  and  $C$ , we find that the first-stage estimates are usually heavily biased due to the lack of strong initial instruments, affecting the quality of Chamberlain instruments in the second stage. To address this issue, we implement a three-step procedure by constructing Chamberlain instruments twice.

Table 1: Demand Estimation with Unobserved Information Factor in the Three-type Monte Carlo Simulation

| param.          | true | Three-Type  |          |       |                 |          |       | Three-Type & RC |          |       |
|-----------------|------|-------------|----------|-------|-----------------|----------|-------|-----------------|----------|-------|
|                 |      | Regular IVs |          |       | Chamberlain IVs |          |       | Chamberlain IVs |          |       |
|                 |      | est.        | st. err. | bias  | est.            | st. err. | bias  | est.            | st. err. | bias  |
| (1)             | (2)  | (3)         | (4)      | (5)   | (6)             | (7)      | (8)   | (9)             | (10)     | (11)  |
| $\beta_0$       | 1    | 1.02        | (0.129)  | 0.02  | 1.00            | (0.108)  | 0.00  | 1.00            | (0.113)  | -0.00 |
| $\beta_x$       | 1    | 1.00        | (0.055)  | -0.00 | 1.00            | (0.053)  | -0.00 | 1.00            | (0.058)  | -0.00 |
| $\alpha$        | 1.5  | 1.50        | (0.046)  | -0.00 | 1.50            | (0.045)  | 0.00  | 1.50            | (0.050)  | 0.00  |
| $\phi_A$        | 0.15 | 0.16        | (0.017)  | 0.01  | 0.15            | (0.005)  | -0.00 | 0.15            | (0.007)  | -0.00 |
| $\phi_B$        | 0.5  | 0.48        | (0.033)  | -0.02 | 0.50            | (0.016)  | -0.00 | 0.50            | (0.021)  | 0.00  |
| $\phi_C$        | 0.35 | 0.34        | (0.045)  | -0.01 | 0.35            | (0.018)  | 0.00  | 0.35            | (0.023)  | 0.00  |
| $\gamma$        | -1.5 | -3.13       | (0.140)  | -1.63 | -1.50           | (0.035)  | 0.00  | -1.48           | (0.051)  | 0.02  |
| $\sigma_\gamma$ | 0.5  |             |          |       |                 |          |       | 0.48            | (0.049)  | -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  $\gamma$  and standard deviation  $\sigma_\gamma$ . Regular instruments are the intercept, product characteristics  $x_{jt}$ , prices  $p_{jt}$ , cost shifter  $z_{jt}$ , and two information variables  $k_{1jt}, k_{2jt}$  with  $k_{1jt}$  observed and  $k_{2jt}$  estimated by a linear factor model. Chamberlain IVs replace the latter two information variables in the instrument set.

the random coefficient on the valuation of the expected attribute  $\gamma$ .

Appendix Table 6 displays the estimation results when the DGP includes two information types,  $B$  and  $C$ , and we include  $A$ ,  $B$ , and  $C$  as candidate information types. The estimated mixture assigns zero weight to type  $A$ , correctly recovering that this type is not present in the DGP.<sup>20</sup>

### 6.3 Simulation of the Moment-Inequality Approach

For ease of computation, we modify the setup in Section 6.1 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

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

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

<sup>20</sup>The regular instruments produce a problematic GMM objective function with multiple local minima, as shown in the left panels of Appendix Figure A3. The objective function becomes well behaved with Chamberlain instruments, as shown in the right panels, which further supports their effectiveness.

Table 2: Moment-Inequality Confidence Sets across Selection Thresholds ( $\tilde{\kappa} = A$ )

|            | $\beta_x$<br>(1) | $\gamma$<br>(2) | # Obs.<br>(3) | # Misclass<br>(4) | # Pts in CS<br>(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                 |

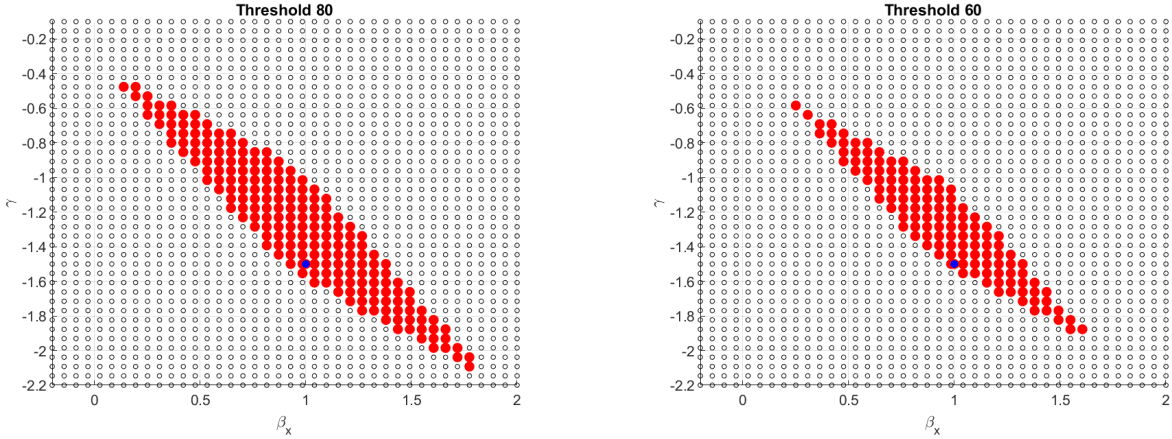
*Notes.* All confidence sets are computed from one simulated dataset of 25,000 observations. The data-generating process includes three information types  $\kappa = A, B, C$ , and we specify  $A$  as the observed extremal type  $\tilde{\kappa}$  in the estimation. The instruments are the intercept,  $x_{jt}$ , and the estimated expectation of the observed type,  $\hat{g}_{Ajt}^e$ . We present results for selection thresholds  $a$  from 0.9 to 0.5. The selected sample sizes are reported in column (3). Among the selected observations, we count the cases in which type  $A$ 's estimated prediction error,  $\hat{e}_{Ajt} \equiv g_{jt} - \hat{g}_{Ajt}^e$ , is not the relevant extremum among  $\{\hat{e}_{\kappa jt}\}_{\kappa \in \{A, B, C\}}$ ; these are reported in column (4). Columns (3) and (4) each aggregate counts across the max and min inequalities (e.g., at the 90th- and 10th-percentile thresholds, we have 2,500 observations for the max side and another 2,500 observations for the min side). In our simulation, all misclassified observations used in estimation come 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  $\beta_x$  ranging from  $-0.2$  to  $2$  with a step of  $0.055$  and  $\gamma$  from  $-2.2$  to  $-0.1$  with a step of  $0.0575$ . Coordinatewise extrema of the nominal 95% joint confidence set are reported in square brackets. The number of grid points  $(\beta_x, \gamma)$  in that confidence set (CS) is reported in column (5).

estimate the moment inequalities in (16) using information type  $A$  as the observed type  $\tilde{\kappa}$  defined in Assumption 9. Consistent with Assumption 8, the candidate grid for  $\gamma$  is restricted to negative values. We select observations where type  $A$ 's prediction errors  $e_{Ajt}$  are above the  $a$ -th quantile  $q_a$  (and conversely below the  $(1 - a)$ -th quantile  $q_{1-a}$ ) of its empirical distribution, with  $a \in \{0.9, 0.8, \dots, 0.5\}$ . The selection conditions are  $Z_{\max jt}^S = \mathbb{1}\{e_{Ajt} \geq q_a\}$  and  $Z_{\min jt}^S = \mathbb{1}\{e_{Ajt} \leq q_{1-a}\}$ .

In this simulation, relaxing the thresholds increases the selected sample size and eventually introduces non-extremal selections according to the estimated prediction errors. We report these classification diagnostics alongside confidence-set sizes to illustrate the practical implications of the validity–power trade-off.

Table 2 illustrates the sample-size and validity considerations behind this trade-off. Lowering the threshold increases the selected sample size and shrinks the confidence set, but eventually introduces misclassifications. Using the estimated prediction errors, no selected observations are classified as non-extremal at  $a = 0.9$  or  $a = 0.8$ . The misclassification rate rises to about 3% at  $a = 0.7$  and exceeds 20% at  $a = 0.5$ , where the joint confidence set excludes the true parameter vector in this simulated sample. Smaller sets based on potentially invalid moments cannot by themselves be interpreted as evidence of greater power. This single-dataset exercise illustrates threshold sensitivity; it does not estimate

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



Notes. This figure plots the nominal 95% confidence regions (in red) for selection thresholds  $a = 0.8$  and  $a = 0.6$ . The blue point indicates the true parameter vector. Lowering the threshold increases the selected sample size and produces a smaller confidence region. At  $a = 0.6$ , around 14% of selected observations are misclassified, but the region still contains the true parameter vector in this simulated sample. Such inclusion does not establish nominal coverage.

coverage probabilities or power.

Figure 1 illustrates this pattern for  $a = 0.8$  and  $a = 0.6$ . Appendix Figure A4 reports the coordinatewise extrema of the confidence sets across all thresholds.

## 7 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 approaches. Second, we revisit the market-level data setting of GRV18 and apply our finite-mixture approach.

### 7.1 What Do Exporters Know?

We revisit DM18, who study the information structure underlying Chilean firms' exporting decisions. Their approach relies on a minimal information set that is assumed to be known to all exporters and uses moment inequalities to obtain identified sets for the parameters governing fixed exporting costs. We construct a finite-mixture analogue of their application that keeps this minimal information set as a common core and makes heterogeneity above it explicit. This allows us to estimate the distribution of information types and quantify the fraction of firms that rely on different information sets when forming predictions about export profitability.

We use the same unbalanced panel studied in DM18, covering  $N = 266$  Chilean firms in the food and chemical sectors over 1996–2005, and firm–destination export decisions to  $J^c = 22$  (chemical) and  $J^f = 34$  (food) destinations. Following DM18, the key uncertain object is the potential export revenue. The information variables highlighted in DM18 include distance to the destination, lagged domestic revenue, lagged aggregate export revenue to the destination, the lagged revenue shifter, and the lagged number of exporters to the destination. To keep our application directly comparable to DM18, we keep their minimal information set  $\mathcal{I}_{ijt}^m = \{dist_j, r_{iht-1}, R_{jt-1}\}$  as a common core and allow types to differ only in whether they additionally use the lagged revenue shifter  $\alpha_{jt-1}$ , the lagged number of exporters  $N_{jt-1}$ , or both. Thus, although the five candidates generate  $2^5 - 1 = 31$  nonempty subsets, our empirical specification deliberately uses the four sets containing  $\mathcal{I}_{ijt}^m$ . This application-specific sublattice preserves comparability with DM18 and is not required by our general framework. Because lagged domestic revenue  $r_{iht-1}$  varies across firms, the resulting prediction  $g_{\kappa,ijt}^e$  remains firm-specific even among firms of the same information type  $\kappa$ .

We estimate a finite-mixture version of the export decision model by maximum likelihood, where the mixture weights  $\phi$  capture the population shares of firms using each information type. In parallel, we report moment-inequality estimates based on the minimal-information benchmark and on our moment inequality approach. Estimation details—including the construction of destination-year revenue shifters, the computation of type-specific predicted revenues, and the implementation of the moment inequalities—are provided in Appendix B.

Figure 2 compares parameter estimates across the perfect-foresight plug-in benchmark, the minimal-information plug-in benchmark, the DM18 inequality estimator, and our information-robust estimators. We confirm that naive plug-in approaches can lead to substantial bias. Our extremal moment-inequality approach delivers an identified set that overlaps with the DM18 set.<sup>21</sup> The finite-mixture confidence intervals overlap with both inequality approaches.

Table 3 reports the estimated information-type shares. We find substantial heterogeneity in firms’ prediction information sets: 15% of firms are consistent with the minimal-information type, 56% incorporate information on the lagged number of exporters, and the type that adds only the lagged revenue shifter receives negligible weight. We also estimate that 29% of firms use the full set of observed information variables. These results

---

<sup>21</sup>Our identified set is wider because we aggregate individual choices into market shares, which reduces the informativeness of the data.

Table 3: Finite Mixture Estimates for Entry Decision in the Chemical Sector

| Parameters | $\sigma$ | $\beta_0$ | $\beta_1$ | $\phi_A$ | $\phi_B$ | $\phi_C$ | $\phi_D$ |
|------------|----------|-----------|-----------|----------|----------|----------|----------|
|            | 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 ( $\sigma, \beta_0, \beta_1$ ) are reported in thousands of year 2000 US\$, and their values scale proportionally with  $\eta$ , 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.

illustrate how the finite-mixture approach recovers not only preference parameters but also the distribution of information types that rationalizes observed exporting behavior.

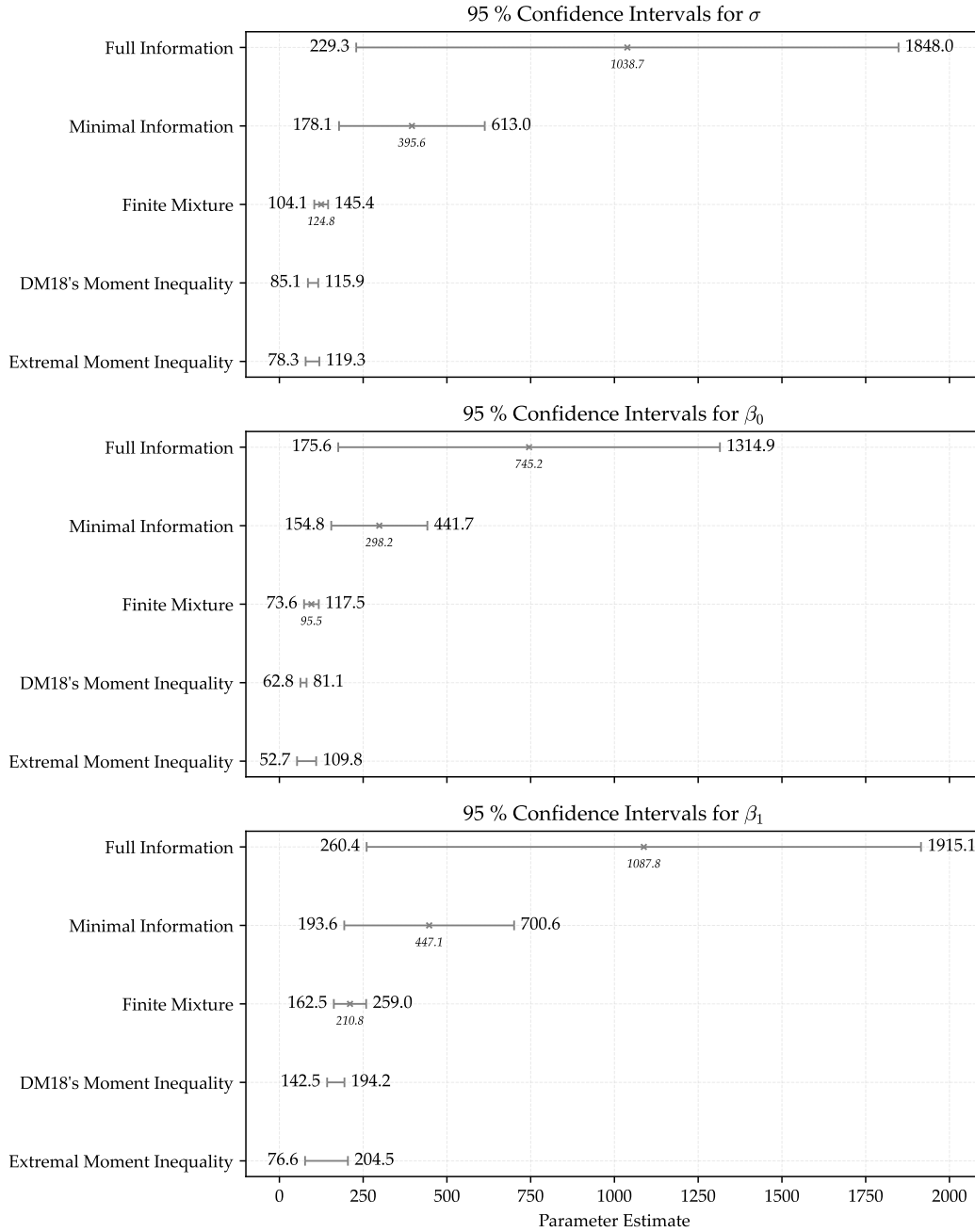
## 7.2 Consumer Valuation of Fuel Costs

A key challenge in the literature on how purchases of energy-consuming durables respond to future operating costs (e.g., Hausman, 1979) is that researchers rarely observe consumers' expectations about future energy expenses at the time of purchase. A common approach is therefore to specify an expectation-formation rule for future expenses and plug the implied expectation into the purchase model. When consumers' actual predictions differ from the researcher's maintained rule, such plug-in approaches are subject to the misspecified-information bias discussed in Section 3.

In this application, we use our framework to allow for heterogeneous prediction information sets when forming expected future expenses and to quantify how alternative prediction types affect inferred valuation of energy. We revisit the estimation framework of GRV18, who study the valuation of automobile fuel costs in the EU market using a demand model with rich preference heterogeneity that combines random coefficients with an empirical mileage distribution. Their estimates suggest that consumers undervalue expected fuel costs at the time of purchase, but the framework does not account for information heterogeneity. We recast the problem in our framework and estimate a finite mixture over observable information types. This delivers fuel-cost valuation estimates consistent with a richer information structure and yields direct estimates of the prevalence of different information types in the data.

**Model.** A consumer  $i$  chooses among vehicle products indexed by model  $j$  and engine variant  $k$  in market  $t$  (defined as a year–country pair). For simplicity, we suppress the

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  $\eta$ , which is set to 5, following DM18. For each parameter  $\sigma$ ,  $\beta_0$ , and  $\beta_1$ , 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. “Full Information” corresponds to the “Perfect Foresight” MLE in DM18, which assumes that firms’ expectations coincide with realized outcomes, i.e.,  $E[r_{ijt} | \mathcal{I}_{ijt}] = r_{ijt}$ . “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. We then 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 structures.

market subscript  $t$ . Expected utility is

$$u_{ijk}^e = x'_{jk} \beta_i^x - \alpha_i (p_{jk} + \gamma g_{ijk}^e) + \xi_{jk} + \epsilon_{ijk},$$

where  $x_{jk}$  is a vector of observed characteristics,  $p_{jk}$  is price,  $g_{ijk}^e$  denotes consumer  $i$ 's expected fuel cost,  $\xi_{jk}$  is an unobserved product attribute, and  $\epsilon_{ijk}$  is a type-I extreme-value error term. The coefficients  $\beta_i^x$  and  $\alpha_i$  capture heterogeneity in tastes for characteristics and in the marginal utility of income.

Expected fuel costs incorporate heterogeneity in mileage. Following GRV18, we write the present discounted value of expected fuel expenditures as

$$g_{ijk}^e = \rho \beta_i^m \mathbb{E}[e_{jk} f_k \mid \mathcal{W}_i],$$

where  $\beta_i^m$  is consumer  $i$ 's annual mileage,  $e_{jk}$  is the fuel-consumption rate (the inverse of fuel efficiency),  $f_k$  is the fuel price, and

$$\rho \equiv \sum_{s=1}^S (1+r)^{-s}$$

is the capitalization factor determined by vehicle lifetime  $S$  and interest rate  $r$ . The parameter  $\gamma$  measures the degree to which consumers internalize future fuel costs at purchase:  $\gamma < 1$  corresponds to under-weighting  $g_{ijk}^e$  relative to the contemporaneous price  $p_{jk}$ .

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  $g_{ijk}^e$ . We assume consumers know their mileage  $\beta_i^m$  and consider the observed fuel-consumption rate  $e_{jk}$  and fuel price  $f_k$  as the information variables used to construct a mixture model of three 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. For each type, the per-mile fuel-cost prediction is  $g_{\kappa,jk}^e = \mathbb{E}[e_{jk} f_k \mid \mathcal{W}_\kappa]$ . This prediction is common across consumers within a market because its information variables vary only at the product–market level, which justifies suppressing  $i$ . Total predicted fuel expenditure nevertheless varies across consumers through mileage and equals  $\rho \beta_i^m g_{\kappa,jk}^e$ . The information type C corresponds to the perfect foresight type that uses both fuel consumption and fuel price to predict fuel costs.<sup>22</sup>

---

<sup>22</sup>We include perfect foresight as an admissible information type, as is common in the fuel-cost literature. We therefore do not impose the timing condition in Extension 2 or interpret type C as its infeasible diagnostic

The diesel indicator has a distinct role in this specification. It enters  $x_{jk}$  as an observed vehicle characteristic, interacted with country dummies, and therefore shifts utility directly; it is not included in any prediction information set  $\mathcal{W}_k$ . Conversely, the fuel-consumption rate and fuel price enter the prediction of fuel costs and do not enter  $x_{jk}$  separately. Although  $f_k$  is the price associated with the fuel used by engine variant  $k$ , information type  $B$  conditions on that price rather than on the diesel indicator. This preserves the distinction in Remark 1 between direct utility characteristics and prediction information variables.

The market share for product  $(j, k)$  is

$$s_{jk} = \sum_{\kappa=1}^{|\mathcal{K}|} \phi_{\kappa} \int \frac{\exp\left(x'_{jk}\beta_i^x - \alpha_i p_{jk} - \alpha_i \gamma \rho \beta_i^m g_{\kappa,jk}^e + \zeta_{jk}\right)}{1 + \sum_{(j',k')} \exp\left(x'_{j'k'}\beta_i^x - \alpha_i p_{j'k'} - \alpha_i \gamma \rho \beta_i^m g_{\kappa,j'k'}^e + \zeta_{j'k'}\right)} dF_{\beta}(\beta_i; \theta), \quad (17)$$

where  $\beta_i \equiv (\beta_i^{x'}, \alpha_i, \gamma \rho \beta_i^m)'$  collects the heterogeneous coefficients entering utility. The component  $\beta_i^x$  follows a parametric distribution indexed by  $\theta$  (means and variance-covariance parameters). The remaining heterogeneity is introduced through empirical draws:  $\alpha_i$  depends on the empirical income distribution, and  $\beta_i^m$  is drawn from the empirical mileage distribution and scaled by  $\gamma \rho$ . The parameter vector is  $\Theta \equiv (\theta', \phi')'$ .

**Estimation.** We estimate the market-share system in (17) using a panel of  $T$  markets. The vector of observed characteristics  $x_{jkt}$  includes horsepower, size and height, an indicator for foreign production, and a diesel indicator interacted with country dummies. Following GRV18, we restrict random coefficients on characteristics to  $\beta_i^x = \bar{\beta}^x + \Sigma^x v_i^x$ , where  $\bar{\beta}^x$  collects mean valuations,  $\Sigma^x$  is diagonal with standard deviations  $\sigma^x$  on the diagonal, and  $v_i^x \sim \mathcal{N}(0, I)$ . The marginal utility of income scales with market income  $y_t$  as  $\alpha_i = \alpha / y_t$ . Individual mileage  $\beta_i^m$  is drawn from the empirical mileage distribution used in GRV18. The parameter vector is

$$\Theta = (\bar{\beta}^{x'}, \sigma^{x'}, \alpha, \gamma \rho, \phi_A, \phi_B, \phi_C)'$$

Finally, unobserved quality is specified as

$$\zeta_{jkt} = \zeta_j + \zeta_t + \tilde{\zeta}_{jkt},$$

where  $\zeta_j$  are model fixed effects and  $\zeta_t$  are market fixed effects (implemented as country-specific linear and quadratic time trends). The model is estimated by GMM using product-

---

type; this application does not provide a formal test of the maintained information structure.

level instruments  $z_{jkt}$  and moment conditions

$$\mathbb{E}[z_{jkt} \tilde{\xi}_{jkt}] = 0,$$

where  $\tilde{\xi}_{jkt}$  is the structural error after partialling out the included fixed effects. Equivalently, the excluded instruments are orthogonal to the unobserved quality component conditional on the included fixed effects.

We approximate the conditional expectation operator  $g_{\kappa,jk}^e = \mathbb{E}[e_{jk}f_k \mid \mathcal{W}_\kappa]$  using a random-forest predictor. This flexibly captures nonlinear relationships between the information variables and expected fuel costs, as illustrated in Appendix Figure A5. In our framework, this first-stage nonparametric prediction provides an empirical estimate of the conditional-expectation mapping and therefore serves as the empirical counterpart to the recoverability clause in Assumption 3(i).

We first estimate the mixture model in Equation (17) without random coefficients using standard two-step GMM. In the first stage, our instruments  $z_{jkt}$  include observed product attributes  $x_{jkt}$ , cost shifters, BLP instruments, and realized fuel costs  $g_{jkt}$ . We then follow Reynaert and Verboven (2014) to construct approximate Chamberlain instruments for the mixture-share parameters  $\phi_\kappa$  and the undervaluation coefficient  $\gamma\rho$  in the second stage.

When allowing for preference heterogeneity, the standard two-step GMM procedure yields imprecise estimates in the mixture model (17). We therefore adopt a one-step estimator that continuously updates the Chamberlain instruments during optimization, following Bourreau et al. (2021). We minimize the GMM objective using Newton’s method with numerically computed gradients.<sup>23</sup>

**Results.** Table 4 reports parameter estimates for the GRV18 specifications in columns (1)–(3) and for the corresponding finite-mixture versions in columns (4)–(6). The key implication is that the data are difficult to reconcile with a perfect-foresight benchmark: the estimated share of the full-information type is essentially zero in the logit and RC Logit I specifications, and remains small in RC Logit II ( $\hat{\phi}_C = 0.10(0.04)$ ). Across specifications, the estimated type distribution therefore assigns little weight to perfect foresight. This result is consistent with external evidence that consumers’ ex-ante and ex-post fuel costs differ substantially (see, for instance, Levinson and Sager, 2023).

The estimated mixture weights reveal substantial heterogeneity in consumers’ expectations about fuel costs. Across specifications, roughly one-half to two-thirds of consumers

---

<sup>23</sup>Accounting for first-stage sampling uncertainty through a bootstrap has little effect on the standard errors for the Logit and RC Logit I specifications and leaves the economic conclusions unchanged. Given the computational cost, we do not implement the analogous bootstrap for RC Logit II.

Table 4: Parameter Estimates for Alternative Demand Models

|                                                   | Table 3 in Grigolon et al. (2018) |                   |                    | Finite Mixture   |                   |                    |
|---------------------------------------------------|-----------------------------------|-------------------|--------------------|------------------|-------------------|--------------------|
|                                                   | Logit<br>(1)                      | RC Logit I<br>(2) | RC Logit II<br>(3) | Logit<br>(4)     | RC Logit I<br>(5) | RC Logit II<br>(6) |
| <i>Panel A. Mean Valuations</i>                   |                                   |                   |                    |                  |                   |                    |
| Price/inc. ( $-\alpha$ )                          | -4.52<br>(0.19)                   | -6.22<br>(0.22)   | -5.33<br>(0.21)    | -2.43<br>(0.21)  | -4.40<br>(0.21)   | -4.97<br>(0.29)    |
| Fuel costs/inc. ( $-\alpha\gamma\rho$ )           | -39.03<br>(1.41)                  | -46.48<br>(0.94)  | -47.11<br>(9.22)   | -12.71<br>(0.33) | -18.41<br>(0.64)  | -11.07<br>(0.46)   |
| Power (kW/100)                                    | 2.28<br>(0.14)                    | 2.60<br>(0.17)    | 0.25<br>(0.61)     | 1.13<br>(0.14)   | 2.23<br>(0.14)    | -1.87<br>(0.05)    |
| Size (cm <sup>2</sup> /10k)                       | 13.25<br>(0.44)                   | 16.69<br>(0.48)   | 16.77<br>(2.02)    | 14.36<br>(0.41)  | 15.76<br>(0.42)   | 14.39<br>(0.52)    |
| Height (cm/100)                                   | 3.00<br>(0.30)                    | 4.45<br>(0.32)    | 5.19<br>(0.33)     | 3.04<br>(0.26)   | 3.85<br>(0.28)    | 5.59<br>(0.24)     |
| Foreign                                           | -0.83<br>(0.02)                   | -0.75<br>(0.02)   | -0.89<br>(0.04)    | -0.94<br>(0.02)  | -0.84<br>(0.02)   | -1.09<br>(0.03)    |
| <i>Panel B. Fractions of Information Types</i>    |                                   |                   |                    |                  |                   |                    |
| Fuel consumption ( $\phi_A$ )                     |                                   |                   |                    | 0.68<br>(0.02)   | 0.66<br>(0.02)    | 0.51<br>(0.02)     |
| Fuel price ( $\phi_B$ )                           |                                   |                   |                    | 0.32<br>(0.03)   | 0.34<br>(0.03)    | 0.39<br>(0.03)     |
| Perfect foresight ( $\phi_C$ )                    |                                   |                   |                    | 0.00<br>(0.04)   | 0.00<br>(0.03)    | 0.10<br>(0.03)     |
| <i>Panel C. Standard Deviations of Valuations</i> |                                   |                   |                    |                  |                   |                    |
| Power (kW/100)                                    |                                   |                   | 1.95<br>(0.25)     |                  |                   | 2.70<br>(0.10)     |
| Size (cm <sup>2</sup> /10k)                       |                                   |                   | 4.31<br>(2.04)     |                  |                   | 6.04<br>(0.62)     |
| Foreign                                           |                                   |                   | 0.49<br>(0.43)     |                  |                   | 1.56<br>(0.06)     |
| Mileage distribution                              | No                                | Yes               | Yes                | No               | Yes               | Yes                |
| <i>Panel D. Valuations of Future Fuel Costs</i>   |                                   |                   |                    |                  |                   |                    |
| Fuel costs/price ( $\gamma\rho$ )                 | 8.63<br>(0.55)                    | 7.47<br>(0.24)    | 8.84<br>(1.77)     | 5.24<br>(0.35)   | 4.19<br>(0.14)    | 2.23<br>(0.19)     |
| Future val. $\gamma$ ( $r = 6\%$ )                | 0.89<br>(0.06)                    | 0.77<br>(0.02)    | 0.91<br>(0.18)     | 0.51<br>(0.03)   | 0.41<br>(0.01)    | 0.22<br>(0.02)     |

*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.

rely primarily on fuel consumption (type *A*), with most of the remainder relying on fuel prices (type *B*). Fuel costs therefore enter choice through heterogeneous prediction rules rather than a common full-information expectation.

Allowing for heterogeneous predictions substantially alters the inferred valuation of future fuel costs. Relative to GRV18, the mixture specification yields markedly lower present-value weights: the implied future-valuation parameter  $\hat{\gamma}$  declines from approximately 0.77–0.91 under the full-information specification to 0.51, 0.43, and 0.22 in columns (4)–(6). These results demonstrate that imposing a plug-in full-information structure attributes cross-sectional variation in beliefs to preferences. As a consequence, standard specifications overstate attention to future operating costs and may generate systematically distorted policy counterfactuals.

### 7.3 Counterfactual policy effects under alternative information structures

GRV18 argue that a fuel tax can reduce aggregate fuel use more than a revenue-equivalent product tax because it targets high-mileage consumers. Table 5 compares their preferred RC Logit II result with two exercises from our model. Because the finite-mixture framework separately identifies preferences and information-type shares, it uniquely allows us to assign every consumer to the perfect-foresight type,  $\tilde{\phi} = (0, 0, 1)$ , while keeping preferences (especially the internalization parameter  $\gamma$ ) estimated under heterogeneous information. GRV18 impose this information structure but do not recover an alternative type distribution. Comparing their result with our all-type-C counterfactual captures the preference-estimation channel; moving from all type C to the estimated mixture changes the information distribution while holding our estimated primitives fixed.

Table 5: Policy effects under alternative information structures

|                                | GRV18<br>$\gamma = 1$ | Our estimates<br>all full information(C)<br>$\gamma = 0.216$ | Our estimates<br>with mixture<br>$\gamma = 0.216$ |
|--------------------------------|-----------------------|--------------------------------------------------------------|---------------------------------------------------|
| Fuel tax                       | −18.13                | −3.86                                                        | −2.61                                             |
| Revenue-equivalent product tax | −12.00                | −14.39                                                       | −18.10                                            |

*Notes.* Entries are percentage changes in total annual fuel use for Germany in 2011. The fuel tax is €0.50 per liter; the product tax is recalibrated to be revenue-equivalent within each scenario. The mixture uses  $\hat{\phi} = (0.51, 0.39, 0.10)$ .

GRV18 find that the fuel tax reduces total fuel use by 18.13%, compared with 12.00% under the product tax. This ordering reverses when applying our preference estimates

and all buyers assigned to type  $C$ , with reductions of 3.86% and 14.39% respectively. The substantially lower estimated valuation of future fuel costs,  $\hat{\gamma} = 0.216$ , makes buyers less responsive to a fuel-price instrument. Evaluating the same model at the estimated mixture reinforces the reversal. Type  $A$ , which accounts for approximately half of consumers, does not condition on fuel prices and, under our maintained updating rule, does not incorporate the fuel-tax increase into expected fuel costs.<sup>24</sup> The fuel-tax effect consequently falls further to 2.61%. In contrast, the product tax enters the purchase price directly and does not require buyers to update fuel-price expectations, creating a larger change in buyer choices. Correcting for biased preference estimates  $\gamma$  and introducing information heterogeneity overturns the quantitative dominance of mileage targeting taxes over product taxes.

## 8 Conclusion

This paper develops an applied framework for estimating discrete choice models when agents form expectations from heterogeneous and unobserved information. Building on the conditional-expectations approach of DM18, the finite-mixture method accommodates individual choices and aggregate market shares, multinomial demand, random coefficients, and endogenous characteristics and estimates preferences and information-type shares. The extremal-inequality method constructs preference bounds from aggregate market shares, an observed prediction rule, and selected observations in which its prediction error is extremal.

The finite-mixture method uses repeated realizations of the uncertain attribute to estimate an augmented predictive basis containing observed predictors and recoverable latent factors. Choice data identify the preference parameters and the population shares associated with the resulting prediction rules. Point identification rests on first-stage recoverability and finite mean-spanning. The identified type-specific demand functions and type shares support counterfactuals that alter product environments or specify alternative information-type distributions. The moment-inequality method uses weaker information about the prediction support and delivers set identification.

The two applications illustrate the role of these tools in empirical work. The exporter estimates are consistent with the bounds in DM18 and recover the prevalence of different revenue-prediction rules. The automobile application finds little support for a common perfect-foresight specification and estimates substantial heterogeneity in fuel-cost predic-

---

<sup>24</sup>The no-update response is an additional counterfactual assumption that treats type- $A$  consumers as not learning or attending to the tax; it does not follow from  $\mathcal{W}_A = \{e\}$  alone.

tions. The framework bridges models of latent information and empirical discrete-choice practice.

## References

- Abaluck, Jason and Abi Adams-Prassl (2021) "What do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses\*," *The Quarterly Journal of Economics*, 136 (3), 1611–1663, 10.1093/qje/qjab008.
- Abaluck, Jason, Giovanni Compiani, and Fan Zhang (2026) "A Method to Estimate Discrete Choice Models That Is Robust to Consumer Search," *Journal of Political Economy*, 134 (7), 1967–2022, 10.1086/740223.
- Agarwal, Nikhil and Paulo Somaini (2026) "Demand Analysis under Latent Choice Constraints," *The Review of Economic Studies*, 93 (4), 2215–2249, 10.1093/restud/rdof093.
- 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](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. (2019) "An Empirical Model of Price Transparency and Markups in Health Care," working paper, National Bureau of Economic Research, [https://websites.umich.edu/~zachb/zbrown\\_empirical\\_model\\_price\\_transparency.pdf](https://websites.umich.edu/~zachb/zbrown_empirical_model_price_transparency.pdf).
- Brown, Zach Y, Christopher Hansman, Jordan Keener, and Andre F Veiga (2023) "Information and Disparities in Health Care Quality: Evidence from GP Choice in England," Working Paper 31033, National Bureau of Economic Research, 10.3386/w31033.
- 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](https://doi.org/10.1016/0304-4076(87)90015-7).
- 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.
- Fox, Jeremy T., Kyoo il Kim, Stephen P. Ryan, and Patrick Bajari (2011) "A simple estimator for the distribution of random coefficients," *Quantitative Economics*, 2 (3), 381–418, 10.3982/QE49.
- 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>.
- 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.
- Gualdani, Cristina and Shruti Sinha (2024) "Identification in Discrete Choice Models with Imperfect Information," *Journal of Econometrics*, 244, 105854, 10.1016/j.jeconom.2024.105854.
- 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>.
- Heiss, Florian, Stephan Hetzenecker, and Maximilian Osterhaus (2022) "Nonparametric estimation of the random coefficients model: An elastic net approach," *Journal of Econometrics*, 229 (2), 299–321, 10.1016/j.jeconom.2020.11.010.
- 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>.
- 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>.
- Vatter, Benjamin (2025) “Quality Disclosure and Regulation: Scoring Design in Medicare Advantage,” *Econometrica*, 93 (3), 959–1001, 10.3982/ECTA21182.
- Wang, Ao (2023) “Sieve BLP: A semi-nonparametric model of demand for differentiated products,” *Journal of Econometrics*, 235 (2), 325–351, <https://doi.org/10.1016/j.jeconom.2022.04.002>.

# Appendices

## Appendix A Proofs

### A.1 Proof of Proposition 1

*Proof.* We show that there exist data generating processes that satisfy Assumptions 1–2 with the same observable distribution. Consider a binary-choice model with  $\beta$  fixed at zero, and suppress individual and market indices. Let  $g \in \{-1, 1\}$  have equal probabilities. The agent observes a signal  $K \in \{-1, 1\}$  and knows its precision parameter  $k \in (0, 1)$ :

$$\Pr(K = g \mid g) = \frac{1+k}{2}.$$

The researcher observes  $(d, X, g)$  but not  $K$  or its precision. Assign  $X$  the same distribution in every structure, independently of  $g$ ,  $K$ , and the taste shocks. The prediction information set is  $\mathcal{W} = \{K\}$ , and Bayes' rule gives

$$g^e = \mathbb{E}[g \mid K] = kK.$$

Thus the signal is informative but imperfect, with  $\text{Var}(g^e) = k^2 > 0$  and  $\mathbb{E}[(g - g^e)^2] = 1 - k^2 > 0$ .

Let the two taste shocks be independent of  $(g, K)$  and i.i.d. Type-I extreme value with scale one. Expected utilities are  $u_1^e = \gamma k K + \epsilon_1$  and  $u_0^e = \epsilon_0$ . Writing  $\Lambda(v) = (1 + \exp(-v))^{-1}$ , the conditional choice probability is  $\Pr(d = 1 \mid K) = \Lambda(\gamma k K)$ . Integrating over the signal conditional on the realized outcome gives

$$\Pr(d = 1 \mid g) = \frac{1}{2} + gk \left[ \Lambda(\gamma k) - \frac{1}{2} \right].$$

We now show that for a set of precision parameter  $k$ , there exists a  $\gamma(k)$  that matches the observed outcomes, thus observed outcomes are not informative of the true  $(k, \gamma)$  pair. Fix a  $r \in (0, 1)$ . For every  $k \in (r, 1)$  ( $k > r$  is implied by the model), there exists a  $\gamma$

$$\gamma(k) = \frac{1}{k} \log \left( \frac{k+r}{k-r} \right) > 0.$$

Substituting in this preference parameter gives  $\Lambda(\gamma(k)k) = (1 + r/k)/2$ , so

$$\Pr(d = 1 \mid g) = \frac{1 + gr}{2}.$$

Thus, for all  $k > r$ , the observed outcomes  $\Pr(d = 1 \mid g)$  is uninformative of the  $(k, \gamma)$  pair. ■

## A.2 Proof of Proposition 2

*Proof.* Let  $\tilde{g}_{jt}^e \equiv (g_{1,jt}^e, \dots, g_{|\mathcal{K}|,jt}^e)' \in \mathbb{R}^{|\mathcal{K}|}$  collect the type-specific predictions for option  $j$  in market  $t$ . Define the type indicator vector  $b_{\kappa(i)} \in \mathbb{R}^{|\mathcal{K}|}$  as the  $\kappa(i)$ -th canonical basis vector.<sup>25</sup> Then  $g_{\kappa(i),jt}^e = b_{\kappa(i)}' \tilde{g}_{jt}^e$ . With individual-specific scale  $\gamma_i$ , define the random-coefficient vector on the prediction component as

$$\tilde{\gamma}_i \equiv \gamma_i b_{\kappa(i)} \in \mathbb{R}^{|\mathcal{K}|}.$$

It follows that

$$\gamma_i g_{\kappa(i),jt}^e = \gamma_i b_{\kappa(i)}' \tilde{g}_{jt}^e = \tilde{g}_{jt}^{e'} \tilde{\gamma}_i.$$

Stack the regular characteristics with heterogeneous tastes  $\tilde{x}_{jt}^{(2)}$  and the prediction vector  $\tilde{g}_{jt}^e$  into

$$x_{jt}^{(2)} \equiv (\tilde{x}_{jt}^{(2)'}, \tilde{g}_{jt}^{e'})', \quad \beta_i \equiv (\beta_i^{(2)'}, \tilde{\gamma}_i')'.$$

Substituting into (11) yields

$$u_{ijt}^e = x_{jt}^{(1)'} \beta - \alpha p_{jt} + x_{jt}^{(2)'} \beta_i + \zeta_{jt} + \epsilon_{ijt},$$

which is a standard random utility model with random coefficients  $\beta_i$ . The information structure implies that the prediction-component coefficient  $\tilde{\gamma}_i$  lies on one of the finitely many directions  $\{b_{\kappa}\}_{\kappa \in \mathcal{K}}$ , scaled by  $\gamma_i$ , completing the isomorphism. ■

## A.3 Proof of Theorem 2

Assumption 3 identifies the candidate prediction functions and requires them to span agents' actual predictions. Taking these functions as known, we apply Berry and Haile (2014)

---

<sup>25</sup>For example, if  $\kappa(i) = 1$  then  $b_{\kappa(i)} = (1, 0, \dots, 0)'$ .

to identify demand, Wang (2023) to identify the induced random-coefficients distribution, and then recover the structural primitives.

Let  $(\beta_i^{(2)}, \gamma_i)$  have distribution  $\tilde{F}_{\beta^{(2)}, \gamma}$  and let  $\kappa(i)$  be independent of these tastes, with probabilities  $\phi$ . Define  $\tilde{\gamma}_i = \gamma_i b_{\kappa(i)}$ , where  $b_\kappa$  is the  $\kappa$ th canonical basis vector. For any Borel set  $\mathcal{B} \subseteq \mathbb{R}^{|\tilde{x}^{(2)}| + |\mathcal{K}|}$ ,

$$F_{\beta^{(2)}, \tilde{\gamma}}(\mathcal{B}) = \sum_{\kappa=1}^{|\mathcal{K}|} \phi_\kappa \int \mathbf{1}\{(\beta_i^{(2)}, \gamma_i b_\kappa) \in \mathcal{B}\} d\tilde{F}_{\beta^{(2)}, \gamma}(\beta_i^{(2)}, \gamma_i).$$

The support of this induced distribution lies on

$$\mathcal{S} = \bigcup_{\kappa=1}^{|\mathcal{K}|} \left( \mathbb{R}^{|\tilde{x}^{(2)}|} \times \{\gamma b_\kappa : \gamma \in \mathbb{R}\} \right).$$

Thus at most one coordinate of  $\tilde{\gamma}_i$  is nonzero. This support restriction is compatible with the mixed-logit model.

### Step 1 (Identification of the demand system).

Following Berry and Haile (2014), condition on a choice set  $\mathcal{J} = \{0, \dots, J\}$  and write  $\chi_t = (x_t^{(1)}, x_t^{(2)}, p_t, \xi_t)$ . Market shares satisfy

$$s_{jt} = \sigma_j(\chi_t) = \Pr\left(\arg \max_{k \in \mathcal{J}} u_{ikt}^e = j \mid \chi_t\right).$$

We distinguish the index used when keeping prices explicit from the full homogeneous utility index:

$$\tilde{\delta}_{jt} = x_{jt}^{(1)'} \beta + \xi_{jt}, \quad \delta_{jt} = \tilde{\delta}_{jt} - \alpha p_{jt}.$$

Keeping prices explicit gives  $s_t = \sigma(\tilde{\delta}_t, p_t; x_t^{(2)})$ ; collecting homogeneous terms gives  $s_t = \bar{\sigma}(\delta_t; X_t^{(2)}, F_{\beta^{(2)}, \tilde{\gamma}})$ , where  $X_t^{(2)}$  has rows  $x_{jt}^{(2)'}$ . The systematic utility of the outside option is zero, so  $u_{i0t}^e = \epsilon_{i0t}$ , and  $\mathbb{E}[\xi_{jt}] = 0$ .<sup>26</sup> We apply the identification conditions of Berry and Haile (2014), summarized below.

### Assumption BH1 (index restriction).

$$F_{u^e}(\cdot \mid \chi_t) = F_{u^e}(\cdot \mid \tilde{\delta}_t, x_t^{(2)}, p_t).$$

### Assumption BH2 (connected substitutes). Goods $(0, 1, \dots, J)$ are connected substitutes in

<sup>26</sup>As in Berry and Haile (2014) and Wang (2023), identification is understood under the maintained index, location, and scale normalizations.

$-p_t$  and in  $\tilde{\delta}_t$ .

**Assumption BH3 (IV exogeneity).**

$$\mathbb{E}[\tilde{\zeta}_{jt} \mid z_t, x_t^{(1)}, x_t^{(2)}] = 0, \quad j = 1, \dots, J.$$

**Assumption BH4 (completeness).** For any measurable  $h(s_t, p_t; x_t^{(2)})$  with finite expectation,

$$\mathbb{E}[h(s_t, p_t; x_t^{(2)}) \mid z_t, x_t^{(1)}, x_t^{(2)}] = 0 \implies h(s_t, p_t; x_t^{(2)}) = 0 \text{ a.s.}$$

BH1–BH2 ensure invertibility. BH3 supplies the exclusion restriction, including the type-specific predictions among the conditioning variables, and BH4 ensures uniqueness of the inverse demand functions.

**Proposition 3.** *Under Proposition 2 and Assumptions BH1–BH4, the structural demand functions  $\bar{\sigma}(\delta; X^{(2)}, F_{\beta^{(2)}, \tilde{\gamma}})$  are identified on their domain.*

*Proof.* This follows from Theorem 1 of Berry and Haile (2014). ■

Identification of demand does not itself identify the mixing distribution, which is the object of Step 2.

### Step 2 (Identification of the random-coefficients distribution).

We apply Theorem 1 of Wang (2023) to the induced coefficient distribution under the following conditions.

**Assumption W1 (logit kernel).** Taste shocks  $(\epsilon_{i0t}, \dots, \epsilon_{iJt})$  are i.i.d. Type-I extreme value with scale one, as maintained in Assumption 1'.

**Assumption W2 (exogenous heterogeneity).** The joint vector  $(\beta_i^{(2)}, \gamma_i, \kappa(i))$  is independent of market characteristics, all prediction regressors, demand shocks, and taste shocks, as maintained in Assumption 7.

**Assumption W3 (regressor rank).** Let  $X_t^{(2)}$  be the matrix with rows  $x_{jt}^{(2)'} = (\tilde{x}_{jt}^{(2)'}, \tilde{\delta}_{jt}^{e'})$ . Its support contains a matrix  $X^*$  of full column rank.

**Assumption W4 (open index domain).** At  $X^{(2)} = X^*$ , the domain on which the structural demand  $\bar{\sigma}(\delta; X^*, F_{\beta^{(2)}, \tilde{\gamma}})$  is identified under the maintained Berry–Haile conditions contains a nonempty open set  $\mathcal{D} \subset \mathbb{R}^J$ .

W3 applies jointly to the ordinary random-coefficient regressors and all prediction regressors. W4 supplies the open variation in the utility-index vector required by Wang’s Condition 1.

**Proposition 4.** *Under Proposition 3 and Assumptions W1–W4, the induced distribution  $F_{\beta^{(2)}, \tilde{\gamma}}$  is globally identified.*

*Proof.* Proposition 3 and W4 identify  $\bar{\sigma}(\delta; X^*, F_{\beta^{(2)}, \tilde{\gamma}})$  on an open set of  $\delta$ , verifying Condition 1 of Wang (2023). W3 verifies Condition 2. Wang’s Theorem 1 therefore globally identifies  $F_{\beta^{(2)}, \tilde{\gamma}}$ . ■

### Step 3 (Identification of the structural primitives).

Given the distribution identified in Step 2, invert the mixed-logit share system to recover  $\delta_t$  uniquely in each market. We now establish identification of our model’s primitives  $(\beta, \alpha, \phi, \tilde{F}_{\beta^{(2)}, \gamma})$  under explicit exogeneity, rank, and common-tastes conditions.<sup>27</sup>

**Identification of  $(\beta, \alpha)$ .** Let  $R_{jt} = (x_{jt}^{(1)'} , -p_{jt})'$  and  $\vartheta = (\beta', \alpha)'$ , so

$$\delta_{jt} = R_{jt}' \vartheta + \xi_{jt}.$$

Choose instrument vectors  $Z_{jt}$  measurable with respect to  $(z_t, x_t^{(1)}, x_t^{(2)})$  for which the following moments are finite. BH3 implies  $m(\vartheta_0) = 0$ , where

$$m(\vartheta) = \mathbb{E} \left[ \sum_{j=1}^J Z_{jt} (\delta_{jt} - R_{jt}' \vartheta) \right].$$

**Assumption 10.** *The population matrix*

$$C = \mathbb{E} \left[ \sum_{j=1}^J Z_{jt} R_{jt}' \right]$$

*exists and has full column rank.*

If  $\vartheta$  and  $\vartheta_0$  both satisfy these moments, subtracting their equations gives  $C(\vartheta - \vartheta_0) = 0$ . Assumption 10 therefore implies  $\vartheta = \vartheta_0$  globally. The rank condition gives global uniqueness because the moments are affine in these coefficients.

**Identification of  $(\phi, \tilde{F}_{\beta^{(2)}, \gamma})$ .** Because  $b_\kappa$  is a canonical basis vector,

$$\gamma_i = \mathbf{1}' \tilde{\gamma}_i.$$

---

<sup>27</sup>Footnote 17 of Wang (2023) notes the linear-IV recovery of homogeneous coefficients. We state the required conditions explicitly and additionally recover the taste distribution and information-type shares.

Hence the joint distribution  $\tilde{F}_{\beta^{(2)}, \gamma}$  is recovered from the identified distribution  $F_{\beta^{(2)}, \tilde{\gamma}}$  by the known mapping  $(b, v) \mapsto (b, \mathbf{1}'v)$ . This includes any mass at  $\gamma_i = 0$  and its joint distribution with  $\beta_i^{(2)}$ . The common-tastes restriction makes this distribution the same across information types.

For each type  $\kappa$ , let

$$E_\kappa = \{v \in \mathbb{R}^{|\mathcal{K}|} : v_\kappa \neq 0, v_\ell = 0 \text{ for all } \ell \neq \kappa\}.$$

Assumption 7 gives  $\Pr(\tilde{\gamma}_i \in E_\kappa) = \phi_\kappa \Pr(\gamma_i \neq 0)$ . Since  $\Pr(\gamma_i \neq 0) > 0$ ,

$$\phi_\kappa = \frac{\Pr(\tilde{\gamma}_i \in E_\kappa)}{\Pr(\tilde{\gamma}_i \neq 0)}.$$

Thus both the common taste distribution and the information-type shares are globally identified.

**Proposition 5.** *Under Assumptions 7, BH3, and 10, the structural primitives  $(\beta, \alpha, \phi, \tilde{F}_{\beta^{(2)}, \gamma})$  are globally identified from  $(\delta, F_{\beta^{(2)}, \tilde{\gamma}})$ .*

Combining the three steps yields the result.

**Theorem 4.** *Under Propositions 2, 3, 4, and 5, the structural primitives  $(\beta, \alpha, \phi, \tilde{F}_{\beta^{(2)}, \gamma})$  in model (11) are globally identified from market-level data on  $(s_t, x_t^{(1)}, \tilde{x}_t^{(2)}, \tilde{g}_t^e, p_t, z_t)$ .*

## A.4 Proof of Theorem 3

This appendix proves Theorem 3. The result is algebraic: it follows from the finite-mixture structure of market shares and does not require the selection and inference assumptions used for implementation in Section 5.

**Setup.** Fix a market  $t$  with outside option 0 and inside options  $j \in \{1, \dots, J\}$ . Let  $\mathcal{K}_t$  denote a finite set of information types. Type  $\kappa \in \mathcal{K}_t$  has logit choice probabilities

$$s_{\kappa j t}(\theta) = \frac{\exp(\delta_{jt}(\beta) + \gamma g_{\kappa, jt}^e)}{1 + \sum_{j'=1}^J \exp(\delta_{j't}(\beta) + \gamma g_{\kappa, j't}^e)}, \quad s_{\kappa 0 t}(\theta) = \frac{1}{1 + \sum_{j'=1}^J \exp(\delta_{j't}(\beta) + \gamma g_{\kappa, j't}^e)}, \quad (18)$$

where  $\theta = (\beta', \gamma)'$  and  $\delta_{jt}(\beta) \equiv X'_{jt}\beta + \xi_{jt}$ .

Let  $\phi_\kappa \geq 0$  denote the population share of type  $\kappa$ , with  $\sum_{\kappa \in \mathcal{K}_t} \phi_\kappa = 1$ . Observed market

shares satisfy the finite-mixture restriction

$$s_{jt} = \sum_{\kappa \in \mathcal{K}_t} \phi_{\kappa} s_{\kappa jt}(\theta), \quad s_{0t} = \sum_{\kappa \in \mathcal{K}_t} \phi_{\kappa} s_{\kappa 0t}(\theta). \quad (19)$$

Assume  $s_{0t} > 0$ .

For each type, the share ratio simplifies to

$$\frac{s_{\kappa jt}(\theta)}{s_{\kappa 0t}(\theta)} = \exp\left(\delta_{jt}(\beta) + \gamma g_{\kappa, jt}^e\right). \quad (20)$$

*Proof of Theorem 3.* Fix  $(j, t)$ . Using (19), write the observed share ratio as

$$\begin{aligned} \frac{s_{jt}}{s_{0t}} &= \frac{\sum_{\kappa \in \mathcal{K}_t} \phi_{\kappa} s_{\kappa jt}(\theta)}{\sum_{\kappa \in \mathcal{K}_t} \phi_{\kappa} s_{\kappa 0t}(\theta)} \\ &= \sum_{\kappa \in \mathcal{K}_t} \underbrace{\frac{\phi_{\kappa} s_{\kappa 0t}(\theta)}{\sum_{\kappa' \in \mathcal{K}_t} \phi_{\kappa'} s_{\kappa' 0t}(\theta)}}_{=: \omega_{\kappa t}(\theta)} \cdot \frac{s_{\kappa jt}(\theta)}{s_{\kappa 0t}(\theta)}. \end{aligned}$$

Because  $s_{0t} > 0$ , the weights  $\omega_{\kappa t}(\theta)$  are well-defined, satisfy  $\omega_{\kappa t}(\theta) \geq 0$  for all  $\kappa$ , and sum to one. Hence  $\frac{s_{jt}}{s_{0t}}$  is a convex combination of the type-specific ratios  $\left\{ \frac{s_{\kappa jt}(\theta)}{s_{\kappa 0t}(\theta)} \right\}_{\kappa \in \mathcal{K}_t}$ , which implies

$$\min_{\kappa \in \mathcal{K}_t} \left\{ \frac{s_{\kappa jt}(\theta)}{s_{\kappa 0t}(\theta)} \right\} \leq \frac{s_{jt}}{s_{0t}} \leq \max_{\kappa \in \mathcal{K}_t} \left\{ \frac{s_{\kappa jt}(\theta)}{s_{\kappa 0t}(\theta)} \right\}. \quad (21)$$

Substituting the logit share-ratio identity (20) into (21) and applying  $\log(\cdot)$  (monotone) yields the population bounds

$$\delta_{jt}(\beta) + \min_{\kappa \in \mathcal{K}_t} \{\gamma g_{\kappa, jt}^e\} \leq \log \frac{s_{jt}}{s_{0t}} \leq \delta_{jt}(\beta) + \max_{\kappa \in \mathcal{K}_t} \{\gamma g_{\kappa, jt}^e\}. \quad (22)$$

Rearranging gives exactly the population inequalities stated in the theorem (equivalently, the conditional moments in (13)–(14) evaluated at the true  $\theta_0$  are nonnegative). In particular, letting  $\theta_0 = (\beta'_0, \gamma_0)'$  denote the true preference parameters generating the mixture shares, the inequalities hold at  $\theta_0$ , which implies  $\theta_0 \in \Theta_0$  by definition of the identified set. ■

## Appendix B Revisiting DM18

This appendix summarizes the model in DM18 and documents our replication and extensions. We first restate the export-decision model and the information variables considered in DM18. We then describe how we (i) implement the finite-mixture specification over information types and estimate it by maximum likelihood, and (ii) implement the extremal moment-inequality approach.

### B.1 Model

Firm  $i$  decides whether to export to destination  $j$  in year  $t$ :

$$d_{ijt} \equiv \mathbb{1}\{\mathbb{E}[\pi_{ijt} \mid \mathcal{I}_{ijt}] \geq 0\},$$

where  $\mathcal{I}_{ijt}$  denotes the information set used by firm  $i$  in year  $t$  to form predictions about the potential export revenue  $r_{ijt}$ . Profits take the form

$$\pi_{ijt} = \eta^{-1}r_{ijt} - f_{ijt},$$

where  $\eta$  is the demand elasticity (calibrated to  $\eta = 5$  in DM18) and  $f_{ijt}$  is a fixed exporting cost, specified as:

$$f_{ijt} = \beta_0 + \beta_1 \text{dist}_j + v_{ijt},$$

where  $\text{dist}_j$  denotes distance to destination  $j$  and  $v_{ijt}$  is a demand shock. The export decision can be written as

$$d_{ijt} \equiv \mathbb{1}\left\{\eta^{-1}\mathbb{E}[r_{ijt} \mid \mathcal{I}_{ijt}] - \beta_0 - \beta_1 \text{dist}_j - v_{ijt} \geq 0\right\}.$$

The demand shock follows a normal distribution  $v_{ijt} \mid (\mathcal{I}_{ijt}, \text{dist}_j) \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ . The probability that firm  $i$  exports to destination  $j$  in year  $t$  can be written as

$$\Pr(d_{ijt} = 1 \mid \mathcal{I}_{ijt}, \text{dist}_j; \theta) = \Phi \left[ \sigma^{-1} \left( \eta^{-1} \mathbb{E}[r_{ijt} \mid \mathcal{I}_{ijt}] - \beta_0 - \beta_1 \text{dist}_j \right) \right], \quad (23)$$

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  $\alpha_{jt}$  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 destination  $j$  in year  $t$ . Under the assumption that  $\mathbb{E}_{jt} [e_{ijt} \mid \mathcal{I}_{ijt}, r_{iht}, f_{ijt}] = 0$ ,  $\alpha_{jt} r_{iht}$  equals the uncertain attribute  $g_{ijt}$  in our model.

DM18 discuss the role of five observed information candidates: 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 destination  $j$ , last year's revenue shifter  $\alpha_{jt-1}$ , and the number of firms  $N_{jt-1}$  that exported to destination  $j$  in year  $t-1$ . They assume all exporters know the first three information variables and construct the minimal information set  $\mathcal{I}_{ijt}^m = \{dist_j, r_{iht-1}, R_{jt-1}\}$ . To construct a finite-mixture analogue of their application, we restrict the support to the interval sublattice

$$\mathcal{S}^{\text{DM18}} = \left\{ \mathcal{I}_{ijt}^m, \mathcal{I}_{ijt}^m \cup \{\alpha_{jt-1}\}, \mathcal{I}_{ijt}^m \cup \{N_{jt-1}\}, \mathcal{I}_{ijt}^m \cup \{\alpha_{jt-1}, N_{jt-1}\} \right\}.$$

Thus all four types contain the DM18 minimal set and differ only in their use of the two additional candidates. The restriction imposes a common lower bound under set inclusion, not a complete ordering of types. It is specific to the application and is not imposed by our general framework. The four information sets in  $\mathcal{S}^{\text{DM18}}$  are indexed by  $\kappa = A, B, C, D$ , respectively.

## B.2 Estimation with the Finite-Mixture Approach

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 candidates  $\mathcal{K}_{ijt} = (dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1}, N_{jt-1})$ .

We separately estimate the export-decision model in each sector by maximum likelihood. In this context, the likelihood function (12) becomes

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

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

$$\Pr(d_{ijt} = 1 | \mathcal{K}_{ijt}, dist_j; \Theta) = \sum_{\kappa \in \mathcal{K}} \phi_{\kappa} \Phi \left[ \sigma^{-1} \left( \eta^{-1} \mathbb{E} [\alpha_{jt} r_{iht} | \mathcal{W}_{\kappa, it}] - \beta_0 - \beta_1 dist_j \right) \right], \quad (24)$$

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

To implement the likelihood function (24), we calculate the type-specific expectations of potential export revenue,  $\mathbb{E} [\alpha_{jt} r_{iht} | \mathcal{W}_{\kappa, it}]$ , using a nonlinear exponential specification following DM18. For instance, the expectation of potential export revenue for information type  $A$ , the minimal information type, is

$$\mathbb{E} [\alpha_{jt} r_{iht} | \mathcal{W}_{A, it}] = \exp (\gamma_{A1} \log(dist_j) + \gamma_{A2} \log(r_{iht-1}) + \gamma_{A3} \log(R_{jt-1})), \quad (25)$$

where we first run a nonlinear 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  $\hat{\gamma}_{A1}, \hat{\gamma}_{A2}, \hat{\gamma}_{A3}$ . We then predict the unobserved expectation as

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

We repeat this calculation for each information type  $\kappa$ . In implementation, we omit an intercept term from this auxiliary prediction regression. When an intercept is included, the fitted values of  $\hat{\mathbb{E}}[\alpha_{jt} r_{iht} | \mathcal{W}_{\kappa, it}]$  are nearly degenerate in our sample (close to zero across types), which makes the resulting type-specific predictions empirically indistinguishable and prevents the mixture model from separating information types. Dropping the intercept restores meaningful variation in predicted expectations across information types, which is the variation exploited in the finite-mixture estimation.

### B.3 Estimation with Moment Inequalities

Following DM18, we estimate the standard deviation  $\sigma > 0$  of the demand shock  $v_{ijt}$ , and we abstract from any unobserved product attribute  $\xi_{jt}$ . We assume that  $v_{ijt}$  follows an EVT1 distribution instead of a normal distribution. This allows us to directly adapt our inequality specification in Equation (16) as follows: Export revenue is a benefit, so this application uses the positive-sign case of Assumption 8. We therefore interchange the maximal- and minimal-error events relative to the cost normalization and label the indicators below by

the upper and lower predicted-revenue bounds they implement.

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

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

with  $Z_{ijt} = (\text{constant}, \text{dist}_j, \mathcal{I}_{\tilde{\kappa},ijt})$  and  $\mathbb{E}[r_{ijt} \mid \mathcal{I}_{\tilde{\kappa},ijt}] = \alpha_{jt} r_{iht} - e_{\tilde{\kappa}jt}$  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 quantile groups for each destination  $j$  in year  $t$ . The group-specific export probabilities are  $s_{gjt} = \frac{\sum_{i \in g} d_{ijt}}{N_g}$ , where  $N_g$  is the number of firms in revenue group  $g$ . 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. Alternatively, we could specify an alternative parameter grid, but we prefer to evaluate the inequalities on the same grid as in DM18.

We define the minimal information type as the observed type  $\tilde{\kappa}$ . Because only about 11% of firms in the sample (chemical sector) are exporters, the estimated group-specific export probabilities  $s_{gjt}$  are typically very small and estimated from relatively sparse cells. This leads to limited variation in the lower tail of the estimated probabilities and, after the log-odds transformation, provides little identifying variation for the min-side inequalities. As a result, in practice the min-side inequalities are much less informative than the max-side inequalities.

To address this issue, we use asymmetric selection thresholds in Assumption 9. Specifically, after estimating the predicted export revenues  $\hat{\mathbb{E}}[r_{ijt} \mid \mathcal{I}_{ijt}^m]$  for the minimal information type  $\mathcal{I}_{ijt}^m$ , we select extreme predictions below the 5th percentile, i.e.,  $Z_{\min jt}^S = \mathbb{1} \left\{ \hat{\mathbb{E}}[r_{ijt} \mid \mathcal{I}_{ijt}^m] \leq q_{0.05} \right\}$ , to construct the min-side moments. We use predictions above the 93rd percentile, i.e.,  $Z_{\max jt}^S = \mathbb{1} \left\{ \hat{\mathbb{E}}[r_{ijt} \mid \mathcal{I}_{ijt}^m] \geq q_{0.93} \right\}$ , to construct the max-side moments. The asymmetric cutoffs reflect the uneven identifying content of the two tails after aggregating individual choices: the more informative max side supports keeping a somewhat larger selected sample, whereas extending the min side mainly adds sparse cells whose inequalities cannot reject points on the DM18 parameter grid. We therefore study threshold sensitivity in Section 6, where predictions for every information type are observed in the data-generating process and both the effective sample size and the

frequency of non-extremal selections can be reported.

## Appendix C Additional Tables

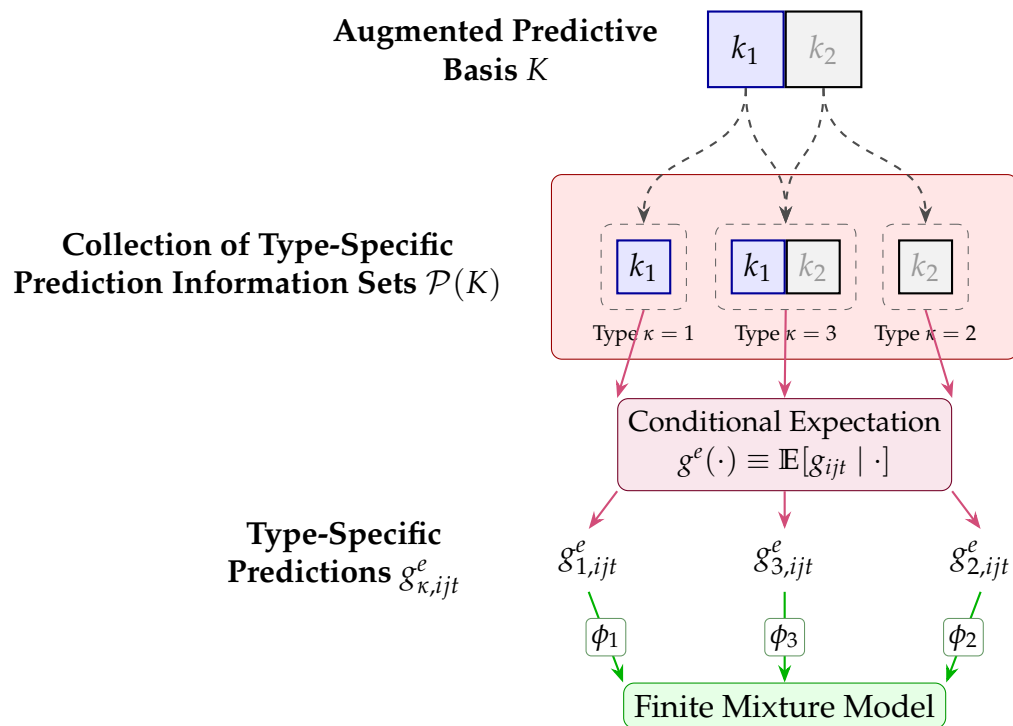
Table 6: Demand Estimation with Unobserved Information Factors in the Two-type ( $B, C$  included) Simulation

| param.    | true | Regular IVs |          |       | Chamberlain IVs |          |       |
|-----------|------|-------------|----------|-------|-----------------|----------|-------|
|           |      | est.        | st. err. | bias  | est.            | st. err. | bias  |
| (1)       | (2)  | (3)         | (4)      | (5)   | (6)             | (7)      | (8)   |
| $\beta_0$ | 1    | 2.58        | (4.160)  | 1.58  | 1.00            | (0.084)  | 0.00  |
| $\beta_x$ | 1    | 1.08        | (1.111)  | 0.08  | 1.00            | (0.053)  | 0.00  |
| $\alpha$  | 1.5  | 1.62        | (1.121)  | 0.12  | 1.50            | (0.045)  | -0.00 |
| $\phi_A$  | 0    | 0.09        | (0.095)  | 0.09  | 2.31E-04        | (0.001)  | 0.00  |
| $\phi_B$  | 0.4  | 0.63        | (0.189)  | 0.23  | 0.40            | (0.018)  | 0.00  |
| $\phi_C$  | 0.6  | 0.29        | (0.157)  | -0.31 | 0.60            | (0.019)  | -0.00 |
| $\gamma$  | -1.5 | -2.93       | (0.867)  | -1.43 | -1.51           | (0.027)  | -0.01 |

*Notes.* All estimates are from 1,000 Monte Carlo simulations and the DGP includes two information types  $\kappa \in \{B, C\}$ . Regular instruments are the intercept, product characteristics  $x_{jt}$ , prices  $p_{jt}$ , cost shifter  $z_{jt}$ , and two information variables  $k_{1jt}, k_{2jt}$  with  $k_{1jt}$  observed and  $k_{2jt}$  estimated by a linear factor model. Chamberlain IVs replace the latter two information variables in the instrument set.

## Appendix D Additional Figures

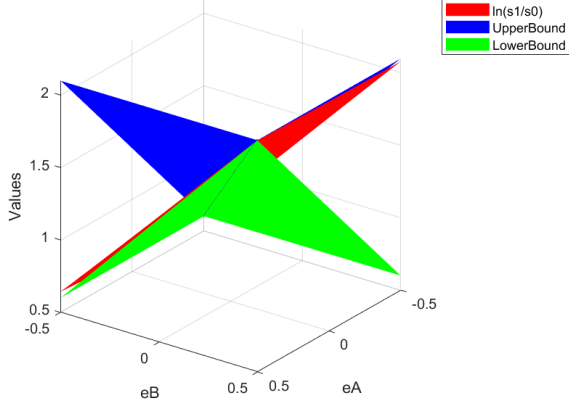
Figure A1: Visual representation of the information structure construction.



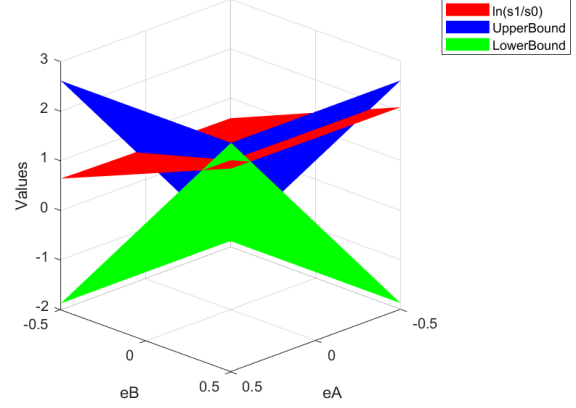
*Notes.* This figure illustrates the construction of type-specific prediction information sets from an augmented predictive basis consisting of an observed predictor  $k_1$  and a latent factor  $k_2$  recovered from repeated realizations of  $g$ . The power set  $\mathcal{P}(K)$  generates distinct information types, which are mapped through the estimated conditional-expectation functions  $g^e(\cdot)$  to form type-specific predictions.

Figure A2: Graphical Illustration of Inequalities

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

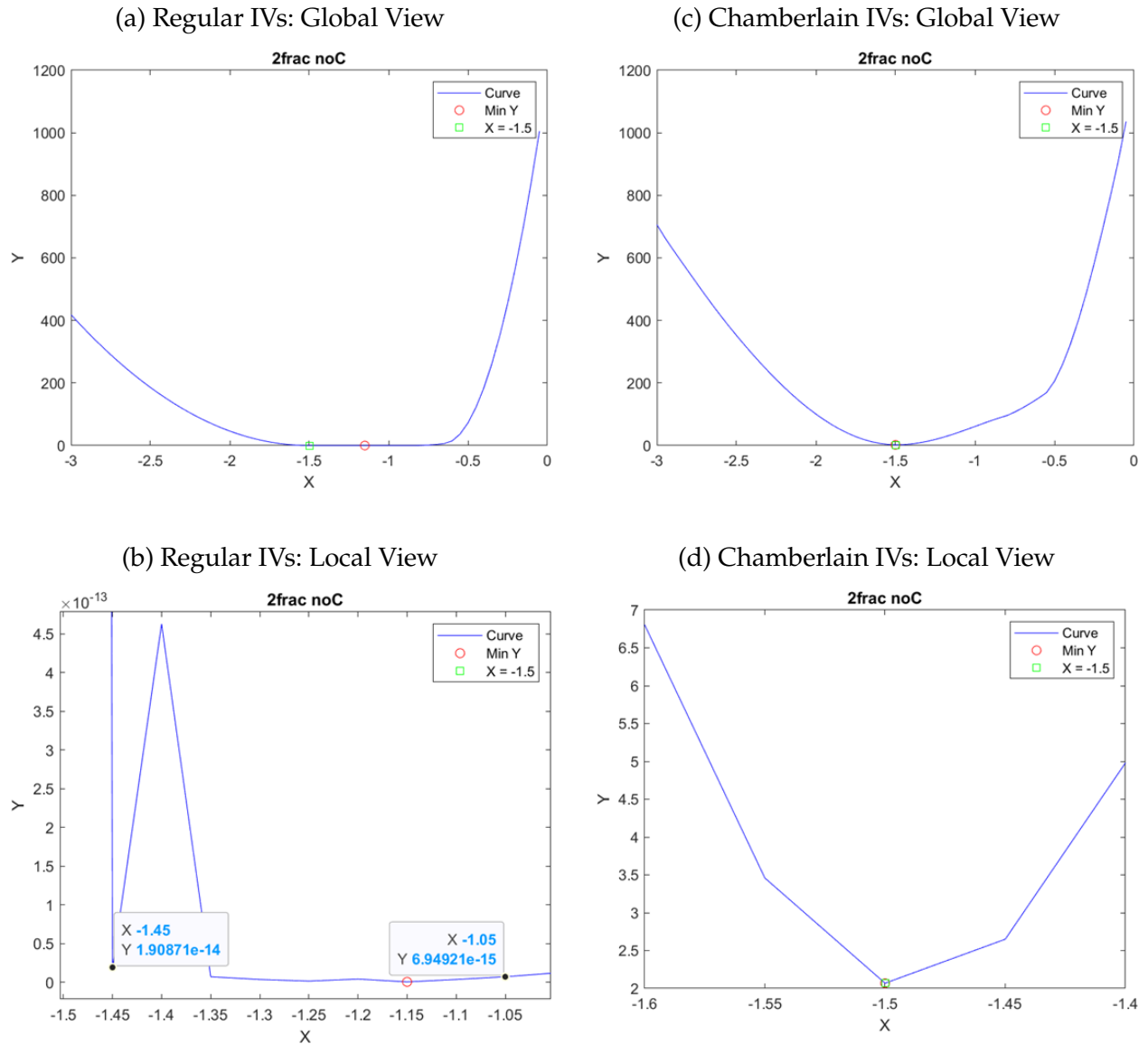


(b) Under False Value of  $\gamma = -4.5$



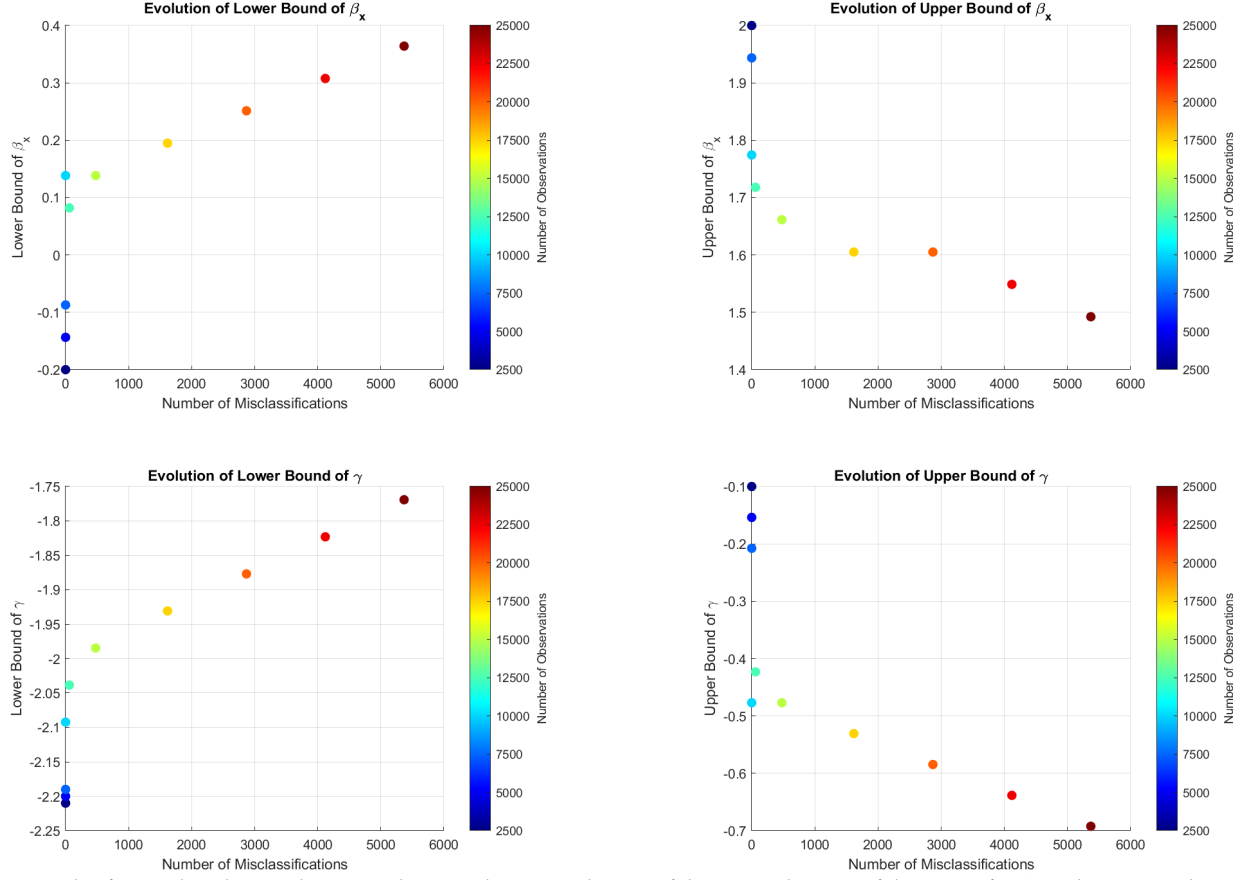
*Notes.* This figure illustrates the identifying content of our inequalities (15) for the parameter  $\gamma$ . The right-hand side of (15) is plotted in blue, while the left-hand side is in green. The observed log market share ratio is plotted in red. Given the true parameter value  $\gamma = -1.5$ , plot A2a 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 an incorrect parameter value  $\gamma = -4.5$ , our inequalities are violated, which rules that value out of the identified set, as depicted in plot A2b.

Figure A3: Comparison of GMM Objective Functions with and without Chamberlain IVs



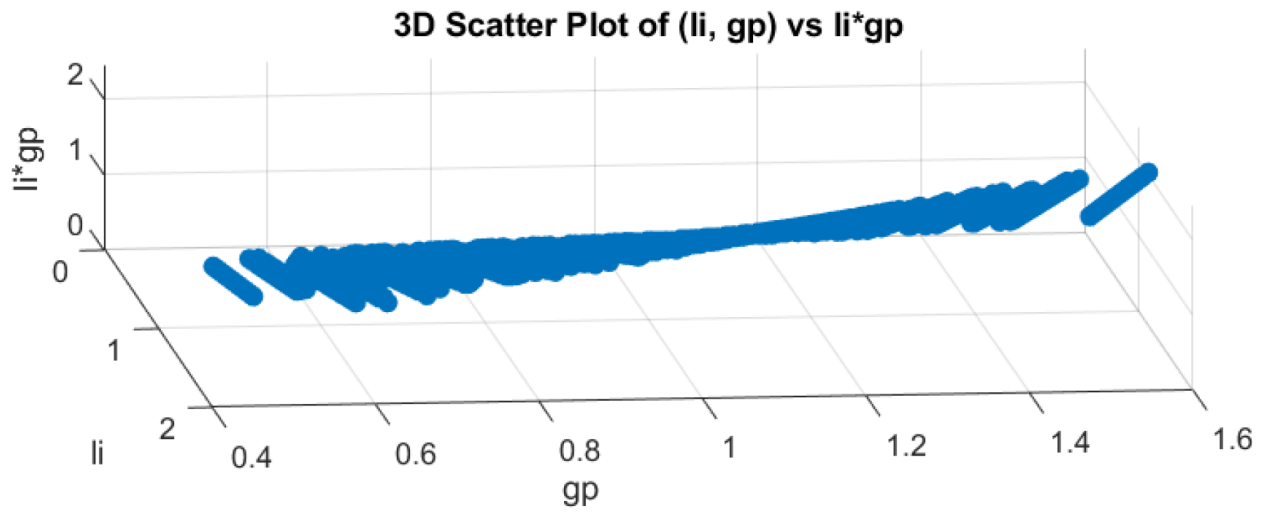
Notes. This figure displays the GMM objective function using two types of instruments: standard IVs in panels (a) and (b) on the left, and Chamberlain IVs in panels (c) and (d) on the right. The top panels (a, c) illustrate the global shape of the objective function, while the bottom panels (b, d) focus on the region where the function attains its minimum. The horizontal axis shows the estimated parameter  $\gamma$ ; 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 instruments, which further supports their effectiveness.

Figure A4: Confidence-Set Projections for  $\beta_x$  and  $\gamma$  across Selection Thresholds ( $\tilde{\kappa} = A$ )



Notes. This figure plots the coordinatewise lower and upper endpoints of the nominal 95% confidence sets for  $\beta_x$  and  $\gamma$  across selection thresholds. The y-axis reports the endpoint value, the x-axis reports the number of misclassified selected observations, and the color indicates the selected sample size. The projections narrow as thresholds are relaxed in this dataset, but their width alone does not establish validity or statistical power.

Figure A5: Nonlinear Relationship between Information Variables and Fuel Costs



Notes. This figure plots the distribution of 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-consumption rates ( $e_{jkt}$ , i.e., "li" on the y-axis). Fuel costs are defined as  $g_{jkt} := e_{jkt} \cdot f_{kt}$ .