Testing and Relaxing Distributional Assumptions on Random Coefficients in Demand Models

Hippolyte Boucher*, Gökçe Gökkoca*, Max Lesellier*

January 2, 2023

Abstract

The BLP demand model for differentiated products is the workhorse model for demand estimation with market-level data. This model uses random coefficients to account for unobserved preference heterogeneity. The shape of the distribution of random coefficients matters greatly for many counterfactual quantities, such as the cost pass-through. In this paper, we develop new econometric tools to test this distribution and improve its estimation under a flexible parametrization. First, we construct new instruments that are designed to detect deviations from the true distribution of random coefficients. Second, we develop a formal moment-based specification test on the distribution of random coefficients. Third, we show that our instruments can be successfully used to estimate a flexible distribution of random coefficients. Finally, we validate our approach with Monte Carlo simulations and an empirical application using data on car purchases in Germany. We also show that these methods extend to the mixed logit demand model with individual-level data.

Keywords: Demand Estimation, Specification Test, Random Coefficients
JEL codes: C35, C36, L13, C52

*Toulouse School of Economics, Hippolyte.Boucher@outlook.com, gokce.gokkoca@tse-fr.eu, max.lesellier@tse-fr.eu; Latest version available at https://github.com/HippolyteBoucher/Selecting-Strong-and-Exogenous-Instruments-via-Structural-Error-Criteria.git

Acknowledgements: The authors thank Christian Bontemps, Nour Meddahi and Mathias Reynaert for their guidance and support as well as Steve Berry, Richard Blundell, Pierre Dubois, Jean-Marie Dufour, Eric Gautier, Cristina Gualdani, Koen Jochmans, Philip Haile, Vishal Kamat, Pascal Lavergne, and Thierry Magnac for helpful comments and discussions. We also thank conference and seminar participants at the TSE Econometrics and the Yale IO workshops as well as the Milan EEA-ESEM, the Brisbane ESAM, the London IAAE and the Tokyo AMES conferences. Finally, we want to thank Kevin Remmy for providing us with a rich data set on car purchases and characteristics in Germany from 2012 to 2018.

1 Introduction

The differentiated product demand model initiated by Berry (1994) and Berry, Levinsohn, and Pakes (1995) has been used in a wide array of empirical studies. It enables researchers to perform demand estimation in markets with differentiated products using either macro-level (market shares) or micro-level (individual purchases) data while allowing for unobserved heterogeneity in preferences as well as price endogeneity. This unobserved heterogeneity in preferences is modeled through the use of random coefficients (RCs) in the utility function. This framework allows researchers to estimate demand functions, price elasticities and counterfactual outcomes. Applications of the BLP model have notably studied the determinants of market power, the welfare effects resulting from a merger or the introduction of a new good and the economic impact of a tax or a subsidy.[1]

The informativeness of the empirical analysis depends on how well the model can reproduce the underlying substitution patterns and approximate the shape of the demand curve, including its slope and curvature. A recent result in Miravete, Seim, and Thurk (2022) shows that the commonly used Gaussian RC on price imposes strong restrictions on the demand's curvature and thus limits the range of the implied pass-through. The degree of pass-through of taxes and costs is central to answering many questions in economics such as the impact of tariffs or a cost shock on consumer welfare. However, estimating a more flexible demand system with a non-Gaussian distribution of random coefficients is challenging. First, there is a clear trade-off between the degree of flexibility one chooses (for instance, going from a Gaussian to a Gaussian mixture) and the precision of the estimates one obtains. Therefore, it is important to be able to test the specification chosen by the researcher on the distribution of the RC (for instance, a Gaussian RC) and quantify the degree of misspecification before potentially moving to a more flexible specification. Second, to precisely estimate a more flexible distribution of RC, the researcher must choose instruments (or equivalently moment conditions) that strongly identify this distribution. The instruments used by the current empirical

[1] The BLP demand model has been widely utilized in numerous applications. A non-exhaustive list of examples includes: Barahona, Otero, Otero, and Kim (2020), Berry et al. (1995), Crawford, Shcherbakov, and Shum (2019), Dubois, Griffith, and O'Connell (2018), Durrmeyer (2022), Grennan (2013), Grigolon, Reynaert, and Verboven (2018), Miller, Sheu, and Weinberg (2021), Miller and Weinberg (2017), Miravete, Moral, and Thurk (2018), Nevo (2000), Petrin (2002), Reynaert (2021).

practice work well with the standard Gaussian RC, but their performance appears to decline as the specification becomes more flexible in the simulation exercises that we perform.

In this paper, we provide novel econometric tools to address these two challenges. In particular, we construct a new set of instruments designed to detect deviations from the true distribution of random coefficients. Building on these instruments, we provide a formal moment-based specification test on the distribution of random coefficients, which can be implemented without having to re-estimate the model under a more flexible parametrization. Our instruments are designed to maximize the power of this test when the distribution of RC is misspecified. We also show how these instruments can strengthen the identifying power of the moment conditions used for estimation, and thus be successful at estimating a flexibly parameterized distribution of RCs. As an example of a flexible parametric distribution, we consider the Gaussian mixture, which can approximate arbitrarily well any continuous distribution on the real line.

This paper consists of three main contributions. First, we construct a new set of instruments that are designed to detect departures from the true distribution of RCs. The intuition we use is the following. Any given distribution of RCs generates a structural error, which, if correctly specified, is mean-independent with respect to a set of exogenous variables. This identifying condition can be transformed into unconditional moments, which can be used to test whether the chosen distribution of RCs is correctly specified. We formally define this test and construct instruments that maximize its power against a fixed alternative. In a first step, we assume that the econometrician knows the fixed alternative and derives an expression for the first-best instrument. We call this instrument the most powerful instrument (MPI) and show that this specific choice of instrument achieves the consistency of the test. In a second step, we provide two feasible approximations of the MPI that can be derived without knowledge of the fixed alternative. We call these feasible MPIs the interval instruments in reference to the way they approximate the MPI.

Second, we consider the case where the researcher wants to test whether the distribution of RCs belongs to a given parametric family. For instance, the researcher may be interested in testing if the random coefficient is normally distributed. This is a composite hypothesis, and we must estimate the unknown parameters of the distribution in a first step. In a second step, we choose instruments to test if the distribution evaluated at the estimated parameters is correctly specified. Here, the interval instruments represent a natural choice of instruments as they are designed to detect deviations from the true distribution of RCs. We study the asymptotic properties of our test when the number of markets, T, goes to infinity and we prove the asymptotic validity of the test under common assumptions. In particular, we account for the statistical uncertainty stemming from the first step estimation, and we control for the magnitude of the approximations that intervene in the estimation of the BLP model. Our asymptotic results complement previous work by Freyberger (2015) on the asymptotic properties of the BLP estimator when the number of markets grows to infinity.

Third, we show that our interval instruments can be successfully used to estimate the model, and particularly so when the distribution of RCs is flexibly parameterized. We do so by exhibiting the connection between the MPI and the classical optimal instruments used for efficient estimation purposes. Specifically, we show that the MPI devoted to testing the specification of the model at the true parameter against any local alternative can be rewritten as a linear combination of the optimal instruments. This relation between the MPI and the optimal instruments helps us understand why the interval instruments, which approximate the MPI, perform so well in our simulations. So far, the literature has exclusively exploited instruments that approximate the optimal instruments (Gandhi and Houde (2019), Reynaert and Verboven (2014)). We refer to these instruments as traditional instruments. These have been shown to work well in the usual Gaussian case. However, our simulations show that their performance declines when we depart from the Gaussian RC.

To evaluate the performance of our test and instruments, we conduct two sets of simulation experiments. First, we compare the performance of the test when using our interval instruments and when using the instruments commonly adopted by practitioners (Gandhi and Houde (2019), Reynaert and Verboven (2014)). We show that the test has the correct empirical size and that the interval instruments significantly outperform the traditional instruments in terms of power under alternative distributions. Second, we evaluate the performance of the interval instruments in estimating the model when the distribution of RC is flexibly parametrized, and follows a Gaussian mixture. We show that our instruments outperform the traditional instruments in terms of the mean squared error. In the case where the RC is a simple Gaussian, the three sets of instruments perform equally well.

Finally, we apply the tools developed in this paper to estimate the demand for cars in Germany from 2012 to 2018. The objective of the empirical exercise is to see how well our instruments perform at estimating a flexible distribution of RCs using a real dataset. Given the importance of price to address most empirical questions, we increase the flexibility of the model by estimating a Gaussian mixture for the RC associated with price. Second, we use our specification test to assess how the degree of misspecification decreases when we increase the flexibility in the distribution of RCs. Third, we use our results to study how the shape of the RC on price can modify important counterfactual quantities such as the pass-through. In particular, our empirical results are consistent with the findings in Miravete et al. (2022).

Related literature Our paper contributes to several strands of the literature. First, it contributes to the literature on the flexible estimation of aggregate demand models for differentiated goods. A few recent papers have proposed non-parametric and semi-parametric methods to estimate aggregate demand functions. Compiani (2018) proposes a non-parametric estimator of the demand functions. If relaxing all the parametric assumptions makes this approach conceptually appealing, it also faces significant theoretical and practical difficulties (more stringent data requirements, large curse of dimensionality, limited scope for counterfactual analysis).[2] Lu, Shi, and Tao (2021) and Wang (2022) propose semi-parametric estimators of the distribution of RCs. These approaches are complementary to ours and the instruments we develop in this paper can be useful to implement their non-parametric IV estimation procedures, which are known to be rather sensitive to the quality of the instruments (Chetverikov and Wilhelm (2017)). Finally, Ho and Pakes (2014), Tebaldi, Torgovitsky, and Yang (2019) suggest deriving bounds directly on the counterfactual quantities.

Our paper also contributes to the literature on the non-parametric identification of the distribution of RCs in demand models (Fox and Gandhi (2011), Fox, il Kim,

[2] In particular, Compiani (2018) relaxes the Type 1 Extreme Value assumption on the taste shock. However, it is not clear how restrictive this assumption is. McFadden and Train (2000) shows that a mixed-logit model with flexibly distributed random coefficients can approximate any discrete choice model derived from random utility maximization. On the other hand, the Type 1 Extreme Value assumption generates massive computational gains, which allows for studying sophisticated markets with many products and many characteristics. Thus, the cost-benefit analysis seems to be largely in favor of the logit specification.

Ryan, and Bajari (2012), Dunker, Hoderlein, and Kaido (2022), Wang (2022), Berry and Haile (2014)). First, we slightly extend the identification result in Wang (2022) to link it directly to the primitives of the model, without assuming that the demand functions are identified. Second, we provide a practical way of constructing moments that feature high identifying power with respect to the distribution of RCs.

Third, we contribute to the literature that focuses on the practical estimation of the BLP model. First, we show that the interval instruments that we construct in this paper can be successfully used to estimate the distribution of random coefficients, and particularly so under of flexible distribution of RCs. This new set of instruments complements instruments commonly used by practitioners: Reynaert and Verboven (2014) and Gandhi and Houde (2019) (see Conlon and Gortmaker (2020) for a review). Moreover, we provide a new parametrization of the model, which facilitates the estimation when the distribution of RCs is a Gaussian mixture. This new parametrization complements previous papers that aim at improving the estimation of the model (Dubé, Fox, and Su (2012), Lee and Seo (2015), Salanié and Wolak (2019)).

Finally, our paper contributes to the literature on the asymptotic properties of the BLP estimator (Armstrong (2016), Berry, Linton, and Pakes (2004), Freyberger (2015), Ketz (2019)). In particular, we prove the asymptotic normality and the consistency of the BLP estimator in the large market framework under less stringent assumptions than the remainder of the literature.

Structure of the paper In Section 2, we recall the baseline BLP model, define the structural error of the model, and provide conditions under which the distribution of RCs is non-parametrically identified. In Section 3, we derive the most powerful instrument and show how it relates to the classical optimal instruments. In Section 4, we construct two feasible approximations of the MPI. In Section 5, we present our specification test and show its asymptotic validity. In Section 6, we conduct Monte Carlo simulations to evaluate the consequences of misspecification on quantities of interest, and gauge the performance of our test and instruments. In Section 7, we apply our new tools to estimate the demand for cars in Germany. We conclude the paper in section 8.

2 Model and identification

2.1 Indirect utility and moment restrictions

Indirect utility We first describe the indirect utility function that induces the observed market shares. Our setting closely follows the one introduced in the seminal paper Berry et al. (1995). There are T markets indexed by $t = 1, ..., T$. There is a continuum of consumers indexed by i. There are J_t market-specific products in market t. Each consumer chooses a product $j \in \{0, 1, ..., J_t\}$ where $j = 0$ corresponds to the outside option. For the sake of exposition and without loss of generality, we will assume throughout our analysis that the number of products is constant across markets ($\forall t,\ J_t = J$). Product j is characterized by a vector of characteristics x_{jt}, which includes the price of the good in most empirical settings. Consumer i derives an indirect utility u_{ijt} from purchasing good $j \in \{0, 1, \dots, J\}$ in market t:

$$u_{ijt} = \underbrace{x'_{1jt}\beta + \xi_{jt}}_{\delta_{jt}} + x'_{2jt}v_i + \varepsilon_{ijt}, \tag{2.1}$$

with the following:

- x_{1jt} is a vector of product characteristics of dimension K_1 associated with product j and for which there is no preference heterogeneity; β represents preferences for x_{1jt};
- ξ_{jt} is an unobserved demand shock on product j in market t;
- $\delta_{jt} \equiv x'_{1jt}\beta + \xi_{jt}$ denotes the mean utility for product j, the part of the utility that is common to all consumers;
- x_{2jt} is a vector of product characteristics of dimension K_2 for which there is preference heterogeneity; v_i is the associated random coefficient that follows a distribution characterized by density f and is independent of all the other variables: $v_i \perp\!\!\!\perp (x_t, \xi_t, \{\varepsilon_{ijt}\}_{j=1,...J})$;
- ε_{ijt} is a preference shock that follows an Extreme Value type I (EV1) distribution independent of all other variables and across i, j, t.

For individual i in market t, the indirect utility from purchasing the outside option is normalized to $u_{i0t} = \varepsilon_{i0t}$. From the random utility functions in (2.1), we can infer the demand functions for each good j in market t denoted $\rho_{jt}(f,\beta)$. Each consumer chooses the product that maximizes his or her utility. Let y_{ijt} equal 1 if individual i chooses good $j = 0, 1, \dots, J$ in market $t = 1, \dots, T$. We have the following:

$$\begin{aligned}
\forall j \neq 0, \quad \rho_{jt}(f,\beta) &\equiv \mathbb{P}_{f,\beta}(y_{ijt} = 1 | x_t, \xi_t) \\
&= \mathbb{P}_{f,\beta}(\text{good } j \text{ is chosen in market } t \text{ by individual } i | x_t, \xi_t) \\
&= \mathbb{P}_{f,\beta}(u_{ijt} > u_{ikt} \ \forall k \neq j | x_t, \xi_t) \\
&= \int_{\mathbb{R}^{K_2}} \frac{\exp\left\{x'_{1jt}\beta + \xi_{jt} + x'_{2jt}v\right\}}{1 + \sum_{k=1}^{J} \exp\left\{x'_{1kt}\beta + \xi_{kt} + x'_{2kt}v\right\}} f(v) dv. \qquad (2.2)
\end{aligned}$$

For the outside option, the demand function is written as follows:

$$\rho_{0t}(f,\beta) = \mathbb{P}_{f,\beta}(y_{i0t} = 1 | x_t, \xi_t) = \int_{\mathbb{R}^{K_2}} \frac{1}{1 + \sum_{k=1}^{J} \exp\left\{x'_{1kt}\beta + \xi_{kt} + x'_{2kt}v\right\}} f(v) dv.$$

Following the EV1 assumption on the idiosyncratic shock on utility, the demand functions take the usual logit form integrated over the distribution of preference heterogeneity. We assume in this paper that the observed market shares are equal to the shares generated by the model above at the true distribution f and the true preference parameter β:

$$\forall j, \quad \forall t, \qquad s_{jt} = \rho_{jt}(f,\beta). \qquad (2.3)$$

Moment restrictions Following the literature, we assume that the unobserved demand shock ξ_{jt} is mean independent of z_{jt}, a set of instrumental variables, namely, $\mathbb{E}[\xi_{jt} | z_{jt}] = 0 \ a.s.$. The set z_{jt} traditionally consists of the exogenous characteristics of all the products on the market as well as cost shifters, which are meant to instrument for price. Indeed, the price of a good is usually considered to be an endogenous variable since it is correlated with the unobserved demand shock ξ_{jt} through the profit maximization problem of firms.[3] To estimate the model, the researcher chooses functions of the instruments z_{jt} to construct a set of unconditional moments. We refer to these functions as estimation instruments and denote them $h_E(z_{jt})$. Likewise, in our analysis,

[3] To deal with the endogeneity of prices, Berry et al. (1995) also suggests using exogenous own-product characteristics as well as exogenous characteristics from other products. The main idea behind the use of these instruments is to take advantage of the correlation between price and exogenous characteristics implied by profit-maximizing firms. To be precise, Berry et al. (1995) suggests using the sum of the characteristics from other products produced by the same firm and the sum of exogenous characteristics from rival firms' products as instruments.

we study the functions of the instruments that are designed to test the specification of the model. We refer to these instruments as testing instruments and we denote them $h_D(z_{jt})$, where D stands for detection.

2.2 Inverse demand function and structural error

Inverse demand function For any given distribution of random coefficients $\tilde{f}$, we define the demand function $\rho \equiv (\rho_1(\cdot), ..., \rho_J(\cdot))$ as the function which maps the vector of mean utilities δ to the vector of market shares generated by the model under $\tilde{f}$:

$$\begin{aligned} \rho(\cdot, x_{2t}, \tilde{f}):\ & \mathbb{R}^J \to [0,1]^J \\ & \delta \mapsto \int_{\mathbb{R}^{K_2}} \frac{\exp\left\{\delta + x'_{2jt}v\right\}}{1 + \sum_{k=1}^J \exp\left\{\delta_k + x'_{2kt}v\right\}} \tilde{f}(v)dv. \end{aligned}$$

Berry (1994) shows by applying Brouwer's fixed point that for any (s_t, x_{2t}) and for any distribution of random coefficients $\tilde{f}$ (even when $\tilde{f}$ is not the true distribution), there exists a unique $\tilde{\delta} \in \mathbb{R}^J$ such that:

$$s_t = \rho(\tilde{\delta}, x_{2t}, \tilde{f}).$$

We define the solution to the previous system of equations as the inverse demand functions: $\rho^{-1}(s_t, x_{2t}, \tilde{f}) = \tilde{\delta}$. Unfortunately, there is no closed form expression for the inverse demand function, which must be recovered numerically.

Structural error From what precedes, we can uniquely define the structural error $\xi_{jt}(\tilde{f}, \tilde{\beta})$ generated by a distribution of random coefficient $\tilde{f}$ and a homogeneous parameter $\tilde{\beta}$:

$$\xi_{jt}(\tilde{f}, \tilde{\beta}) = \rho_j^{-1}(s_t, x_{2t}, \tilde{f}) - x'_{1jt}\tilde{\beta}. \tag{2.4}$$

The non-linear nature of the model is captured by the inverse demand function which enters the expression of the structural error. The absence of an analytical formula for the inverse demand implies that there is no closed form expression for the structural error, which complicates the estimation of the BLP demand model. If we consider a parametric family of distributions $\tilde{\mathcal{F}} = \{\tilde{f}(\cdot|\tilde{\lambda}) : \tilde{\lambda} \in \tilde{\Lambda}\}$, then the structural error generated by a specific element in $\tilde{f}(\cdot|\tilde{\lambda}) \in \tilde{\mathcal{F}}$ and $\tilde{\beta}$ is defined as follows:

$$\xi_{jt}(\tilde{f}(\cdot|\tilde{\lambda}), \tilde{\beta}) = \rho_j^{-1}(s_t, x_{2t}, \tilde{f}(\cdot|\tilde{\lambda})) - x'_{1jt}\tilde{\beta}.$$

2.3 Non-parametric identification

The main objective of this paper is to provide tools to test the specification on the distribution of random coefficients and to improve its estimation under a flexible specification. A natural first step is to study the conditions under which this distribution is non-parametrically identified. The identification of random coefficients in multinomial choice models has been studied extensively in the literature (Allen and Rehbeck (2020), Berry and Haile (2014), Dunker et al. (2022), Fox and Gandhi (2011), Fox et al. (2012), Wang (2022)). We summarize some of these findings in Appendix C.1. In this Section, we build on an important identification result in Wang (2022) to recover a set of sufficient identifying conditions directly on the primitives of the model. We also show that the identification result holds with a less stringent exogeneity assumption than in Wang (2022).

In contrast to the rest of the literature, Wang (2022) adopts all the parametric assumptions in the standard BLP model and looks for a set of sufficient restrictions under which the identification of the demand functions implies the identification of the distribution of random coefficients. This approach allows him to obtain conditions that are less stringent than the rest of the literature. In particular, Wang (2022) makes no special regressor assumption, no full support assumption, and no continuity assumption on the covariates. Specifically, he shows that if the demand functions $\rho = (\rho_1, ..., \rho_J)$ are identified on an open set of $\mathbb{R}^J$, then the distribution of random coefficients is identified.[4] His proof exploits the real analytic property of the demand functions.[5] In this paper, we build on this injectivity result to find sufficient identifying conditions directly on the primitives of the model (without assuming identification of the demand functions). We also show using a random permutation of the indices that we only require the demand shock ξ_{jt} to be mean independent of the instrumental variables z_{jt} across products, but we do not require this to hold for each product j taken separately. Formally, we only require $\mathbb{E}[\xi_{jt}|z_{jt}] = 0$ $a.s.$ and not $\mathbb{E}[\xi_{jt}|z_{jt}] = 0$ $a.s.$ for all product j as previously. This is less restrictive, as demand shocks can now be on average non-zero for certain products and account for unobserved quality inherent to each product.

[4] Identification of demand functions can be achieved using Theorem 1 in Berry and Haile (2014).

[5] In particular, the real analytic property yields that the local identification of ρ on $\mathcal{D} \subset \mathbb{R}^J$ implies the identification of ρ on $\mathbb{R}^J$ From the global identification of ρ, he is then able to show that the random coefficients' distribution is identified under a simple rank condition on x_{2t}.

Let us formally state the assumptions that we impose to recover the point identification of (f, β).

Assumption A

(i) Strict exogeneity: $\mathbb{E}[\xi_{jt}|z_{jt}] = 0$ *a.s.;*
(ii) Completeness: for any measurable function g *such that* $\mathbb{E}[|g(s_t, x_t)|] < \infty$*, if* $\mathbb{E}[g(s_t, x_t)|z_{jt}] = 0$ *a.s., then* $g(s_t, x_t) = 0$ *a.s.;*
(iii) The distribution of the data $(s_t, x_{2t}, x_{1t}, z_t)$ *is fully observed by the econometrician and market shares* s_t *are generated by the demand model defined in section 2.1 by equations (2.1) and (2.3);*
(iv) Detectable difference in distributions: we say f *and* $\tilde{f}$ *differ (and write* $f \neq \tilde{f}$*) if there exists* $\bar{v} \in \mathbb{R}^{K_2}$ *such that* $F(\bar{v}) \neq \tilde{F}(\bar{v})$*;*
(v) Let $x_t = (x_{1t}, x_{2t})$ *then* x_t *is such that* $\mathbb{P}(x_t' x_t$ *is positive definite*$) > 0 \ \ \forall t$*;*
(vi) There exists $\bar{x}_t \in \mathcal{X}$ *and an open set* $\mathcal{D} \subset \mathbb{R}^J$ *such that* $\delta_t = \bar{x}_{1t}\beta_0 + \xi_t$ *varies on* $\mathcal{D}$ *a.s..*

In A(i), we assume that the instruments are strictly exogenous. Assumption A(ii) is a completeness assumption that states that the instruments are strongly relevant with respect to (s_t, x_t). This assumption is typical of semiparametric or nonparametric IV models and is equivalent to a full rank assumption in a linear IV model. Intuitively, it means that if the inverse demands are different almost surely, then the instruments will be able to detect the difference. The completeness assumption is a strong assumption that has been widely used in this literature (Berry and Haile (2014), Dunker et al. (2022), Wang (2022)). Assumption A(v) is a standard rank condition. Assumption A(vi) is meant to ensure that there is enough variation in δ_t to apply the injectivity result in Wang (2022). This assumption indicates that there needs to be sufficient variation in product characteristics across markets in the data to identify f. In practice, product characteristics are very similar from one market to the other and may not yield sufficient variation. A judicious solution is to create inter-market variation by interacting product characteristics with demographic variables characterizing each market. Let us now state our formal identification result.

Proposition 2.1

Under Assumption A, the distribution of random coefficients f *and the homogeneous* *preference parameters β are non-parametrically identified:*

$$(\tilde{f},\tilde{\beta})=(f,\beta) \iff \quad \mathbb{E}[\xi_{jt}(\tilde{f},\tilde{\beta})|z_{jt}]=\mathbb{E}\left[\rho_j^{-1}(s_t,x_{2t},\tilde{f})-x'_{1jt}\tilde{\beta}\middle|z_{jt}\right]=0 \quad a.s..$$

The proof is in Appendix B.1. The identification result above entails that under some fairly weak conditions and in the presence of instruments that generate sufficient variation in the product characteristics, the observed data identifies the distribution of random coefficients non-parametrically. Formally, the model is at the true pair (f,β) if and only if the associated structural error is mean independent of the instrumental variables z_{jt}. We use this identification result to show the consistency of our test under a specific choice of instruments that we will characterize thereafter.

3 Detecting misspecification: the most powerful instrument

The aim of this section is to recover the instrument with the greatest ability to detect misspecification in the distribution of RC. To do so, we consider a setting in which the econometrician wants to test a simple hypothesis of the form $\overline{H}_0:(f,\beta)=(f_0,\beta_0)$. The upper bar is used to stress the fact that $\overline{H}_0$ is a simple hypothesis, in contrast to the composite hypothesis $H_0: f\in\mathcal{F}_0$ that we study in section 5. Our approach builds on a simple intuition: if the model under $\overline{H}_0$ is misspecified, then the structural error will depart from the true demand shock ξ_{jt}, and our goal is to find the best instrument to pin down this deviation. We proceed as follows. First, we introduce a moment-based test for $\overline{H}_0$ and we show its asymptotic validity. Next, we derive an analytical expression for the instrument that maximizes the power of our test against a fixed alternative $\overline{H}_a:(f,\beta)=(f_a,\beta_a)$. We call this instrument the most powerful instrument (MPI) and we show how it relates to the classical optimal instruments, derived for efficient estimation purposes. In Section 4, we provide two feasible approximations of the MPI, which have the critical property of being invariant with respect to the alternative $\overline{H}_a$.

3.1 A moment-based test

We want to test $\overline{H}_0:(f,\beta)=(f_0,\beta_0)$ against $H_a:(f,\beta)\neq(f_0,\beta_0)$. For any set of testing instruments $h_D(z_{jt})$, we have the following implication:

$$\overline{H}_0:(f,\beta)=(f_0,\beta_0) \implies \overline{H}'_0: \quad \mathbb{E}[h_D(z_{jt})\xi_{jt}(f_0,\beta_0)]=0.$$

We propose to test $\overline{H}_0$ indirectly through its implication $\overline{H}_0'$, which is a set of unconditional moment conditions. We test $\overline{H}_0'$ with a moment-based test. Our test statistic writes as follows:

$$S_T(h_D, f_0, \beta_0) = TJ\left(\frac{1}{TJ}\sum_{j,t}\xi_{jt}(f_0,\beta_0)h_D(z_{jt})\right)'\hat{\Omega}_0^{-1}\left(\frac{1}{TJ}\sum_{j,t}\xi_{jt}(f_0,\beta_0)h_D(z_{jt})\right), \tag{3.5}$$

with $\hat{\Omega}_0$ a consistent estimator of Ω_0 the asymptotic variance-covariance matrix of

$$\frac{1}{\sqrt{TJ}}\sum_{j,t}h_D(z_{jt})\xi_{jt}(f_0,\beta_0)$$

which is $\Omega_0 = \mathbb{E}[\xi_{jt}^2(f_0,\beta_0)h_D(z_{jt})h_D(z_{jt})']$. We study the asymptotic properties of our test as the number of markets, T, goes to infinity. As the focus of this section is on the construction of the most powerful instrument, we postpone the treatment of the specific challenges implied by parameter uncertainty (i.e. when β_0 and f_0 must be estimated beforehand) and by the numerical approximations involved in the derivation of the structural error (in practice, the researcher derives a numerical approximation of $\xi_{jt}(f_0,\beta_0)$) to Section 5. Additionally, to keep the results as simple as possible while retaining the key intuitions, we assume independence of the demand shocks in a given market conditional on z_{jt}. This last assumption is relaxed in the proofs in Appendix B.2 and in section 5.

Proposition 3.1

Assume that (s_t, x_t, z_t) are i.i.d. across markets and consistent with the probability model defined by equations (2.1), (2.2) and (2.3) evaluated at (f,β), $\mathbb{E}[\|\xi_{jt}(f_0,\beta_0)h_D(z_{jt})\|^2] < +\infty$, Ω_0 has full rank, and, for $k \neq j$, $\xi_{jt} \perp\!\!\!\perp \xi_{kt}|z_t$. We have the following:

- *under $\overline{H}_0 : (f,\beta) = (f_0,\beta_0)$,* $\qquad S_T(h_D, f_0, \beta_0) \xrightarrow[T\to+\infty]{d} \chi^2_{|h_D|_0}$,
- *under H_a' :* $\mathbb{E}\left[h_D(z_{jt})\xi_{jt}(f_0,\beta_0)\right] \neq 0, \qquad \forall q \in \mathbb{R}^+, \quad \mathbb{P}(S_T(h_D, f_0, \beta_0) > q) \xrightarrow[T\to+\infty]{} 1$,

with $|\cdot|_0$ being the counting norm.

The previous proposition indicates that as long as the testing instruments are functions of z_{jt}, our test procedure is asymptotically valid for $\overline{H}_0$. We are testing $\overline{H}_0$ by virtue of its implication $\overline{H}_0' : \mathbb{E}\left[h_D(z_{jt})\xi_{jt}(f_0, \beta_0)\right] = 0$ and, as a consequence, the power properties of our test hinge critically on the choice of the testing instruments $h_D(z_{jt})$. This is the focus of the next subsection.

3.2 The most powerful instrument (MPI)

The choice of testing instruments $h_D(z_{jt})$ is key to maximizing the rejection rate of $\overline{H}_0$ under any alternative $H_a : (f, \beta) \neq (f_0, \beta_0)$. To guide our choice of instruments, we first derive the instrument that maximizes the power of our test when the econometrician tests $\overline{H}_0$ against a fixed alternative $\overline{H}_a : (f, \beta) = (f_a, \beta_a) \neq (f_0, \beta_0)$. We refer to this instrument as the most powerful instrument (MPI). In practice, the researcher is often reluctant to fix the alternative. However, the MPI represents a useful first-best solution for which we provide two feasible approximations in section 4.

Derivation of the most powerful instrument To construct the MPI, we use the following decomposition of the structural error generated under $\overline{H}_a$:

$$\xi_{jt}(f_0, \beta_0) = \underbrace{\xi_{jt}(f_a, \beta_a)}_{\text{true error under } \overline{H}_a} + \underbrace{\xi_{jt}(f_0, \beta_0) - \xi_{jt}(f_a, \beta_a)}_{\Delta_{0,a}^{\xi_{jt}}},$$

with $\Delta_{0,a}^{\xi_{jt}}$ being the correction term due to misspecification under the alternative $\overline{H}_a$. Our goal is to compare the ability of our test for different candidates $h_D(z_{jt})$, to reject $\overline{H}_0$ under $\overline{H}_a$.

The literature offers many ways to compare the power of competing tests (see Gourieroux and Monfort (1995) for a comprehensive review). First, we distinguish between exact and approximate methods. Exact methods rely on the exact distribution of the test statistic (under $\overline{H}_0$) and allow for comparison in finite sample while asymptotic methods exploit the asymptotic distribution of the test statistic and are informative in larger samples. In our case, the exact distribution of our test is unknown. Thus, we rely on asymptotic methods, which is the most common case in the literature. Second, we divide the methods into local and non-local methods. In parametric tests, local strategies are based on the analysis of the power properties of competing tests under a sequence of local alternatives θ_T which converges to θ_0 at a given rate (usually $\frac{1}{\sqrt{T}}$). The econometrician can compare two competing tests by means of their power functions (or more precisely, the limits of these power functions when sample sizes go to $+\infty$). This is called the direct approach. The dual approach, which is known as Pitman's relative efficiency, consists of comparing the rates at which the minimal number of observations must increase to ensure a given level of power. The approach we favor in this paper is the non-local approach developed in Bahadur (1960). Here, the econometrician chooses the test with the smallest level α needed to attain a given power against a fixed alternative and for a given number of observations. In other words, the econometrician chooses the test that minimizes the risk of type I error ceteris paribus.

There are several reasons to favor Bahadur's non-local approach. First, it is better suited for the testing problem we study in this paper. The comparison criterion, known as the asymptotic slope of the test, is in our case straightforward to derive, whereas it is not clear how one should derive Pitman's efficiency criterion when the test concerns non-parametric objects such as distributions. Moreover, we study the properties of our test against a fixed alternative $\overline{H}_a : (f, \beta) = (f_a, \beta_a)$ as in Bahadur's case, which is not necessarily local. Finally, the literature has highlighted many limitations of the local approach. Local criteria are often unable to discriminate between tests even when these tests lead to different decisions (see Silvey (1959)). In addition, as shown in Dufour and King (1991), a locally optimal test in a neighborhood of H_0 may perform very poorly away from H_0.

Let us now present the intuition for Bahadur's comparison approach. From Section 3.1, we have:

$$\text{Under } \overline{H}_0: \qquad S_T \equiv S_T(h_D, f_0, \beta_0) \xrightarrow{d} S \qquad \text{with } \; S = \chi^2_{|h_D|_0}.$$

Following the same notations as in Gourieroux and Monfort (1995), we denote:

$$\Lambda(s) = \mathbb{P}_{\overline{H}_0}(S \geq s).$$

The critical value is usually derived using the asymptotic distribution of the test statistic under H_0. The approximate critical region at a given level α is then given by:

$$CR_\alpha = \{S_T \geq \Lambda^{-1}(\alpha)\} = \{\Lambda(S_T) \leq \alpha\}.$$

The main idea in Bahadur's approach entails deriving the level of the test if one takes the value of the test statistic as the critical value (this is also known as the p-value). Namely:

$$\alpha_T = \Lambda(S_T).$$

Bahadur suggests preferring the test that displays the lowest level α_T at least asymptotically. A formal analysis of the asymptotic behavior of α_T shows that it is better to consider the limit of a transformation of α_T than the limit of α_T directly. This gives rise to the concept of the approximate slope of the test.

Definition 1 (Asymptotic slope of the test)

(i) $K_T = -\frac{2}{T}\log(\Lambda(S_T))$ is the approximate slope of the test,

(ii) Under $\overline{H}_a$: *plim* $K_T = c(f_a, \beta_a)$ is the asymptotic slope of the test,

with *plim*, the limit in probability when $T \to +\infty$.

Under the alternative $\overline{H}_a : (f, \beta) = (f_a, \beta_a)$, consider two sequences of tests based on S_T^1 and S_T^2 with asymptotic slopes $c^1(f_a, \beta_a)$ and $c^2(f_a, \beta_a)$ respectively. The test based on S_T^1 is asymptotically preferred to the test based on S_T^2 in Bahadur's sense if and only if $c^1(f_a, \beta_a) > c^2(f_a, \beta_a)$. To derive the asymptotic slopes of our test, we apply an important result in Geweke (1981), which states that if under H_0: $S_T \xrightarrow[T\to+\infty]{d} \chi^2_q$ (with any $q \in \mathbb{N}^*$), then $\frac{1}{T}S_T \xrightarrow{a.s.} c(f_a, \beta_a)$ (when the limit exists). In our test, the limiting distribution is chi-squared. Thus, the asymptotic slope of our test with instrument $h_D(z_{jt})$ writes:

$$c_{h_D}(f_a, \beta_a) = plim\ \frac{1}{T}S_T(h_D, f_0, \beta_0) = J\mathbb{E}\left[\xi_{jt}(f_0, \beta_0)h_D(z_{jt})\right]' \Omega_0^{-1}\mathbb{E}\left[\xi_{jt}(f_0, \beta_0)h_D(z_{jt})\right].$$

Let us note that the asymptotic slope can also be interpreted as a measure of the speed of divergence of the test statistic in terms of population moments, i.e. speed of divergence $\approx T \times c_{h_D}(f_a, \beta_a)$. In the next proposition, we derive an analytical expression for the instrument that maximizes the slope of the test.

Proposition 3.2 (Most powerful instrument)
Let $\mathcal{H}$ be the set of measurable vectorial functions of z_{jt}. Under any fixed alternative $\overline{H}_a : (f, \beta) = (f_a, \beta_a)$, we have the following:

$$\left(\mathbb{E}\left[\xi_{jt}(f_0, \beta_0)^2 | z_t\right]\right)^{-1}\mathbb{E}[\Delta_{0,a}^{\xi_{jt}} | z_{jt}] \in \underset{h_D \in \mathcal{H}}{argmax}\ \ c_{h_D}(f_a, \beta_a).$$

The proof is given in Appendix B.2. The MPI equals the conditional expectation of the correction term $\Delta_{0,a}^{\xi_{jt}}$ divided by a conditional variance term $\mathbb{E}\left[\xi_{jt}(f_0, \beta_0)^2 | z_{jt}\right]$. For exposition purposes, we drop the conditional variance term in the subsequent analysis and take the homoskedastic MPI $h_D^*(z_{jt}) = \mathbb{E}[\Delta_{0,a}^{\xi_{jt}} | z_{jt}]$ as the reference MPI.[6] Methods have been proposed to estimate the conditional variance term non-parametrically and could be adapted to our case. However, it is well known that conditional variance, which also appears in the formulation of the optimal instruments, is difficult to model and estimate in practice. In the BLP framework, the large dimension of z_{jt} makes the exercise even more difficult. Hence, researchers typically ignore this term or impose a restrictive and ad-hoc structure on the form that it can take (for instance, Reynaert and Verboven (2014)'s approximation of the optimal instruments in the BLP model ignores the variance term). The homoskedastic MPI, $h_D^*(z_{jt})$, features other appealing properties including (i) consistency of the associated test and (ii) maximizing correlation with the structural error under the alternative.[7] For simplicity, in what follows, we refer to the homoskedastic MPI as the MPI.

(i) Consistency By setting h_D equal to h_D^*, our moment-based test becomes consistent against any fixed alternative $\overline{H}_a : (f, \beta) = (f_a, \beta_a) \neq (f_0, \beta_0)$. Namely, we have the following result:

Proposition 3.3 (Consistency of the test with the MPI)
Under Assumption A and the same assumptions as in Proposition 3.1, we have:

$$\overline{H}_a : (f, \beta) = (f_a, \beta_a) \neq (f_0, \beta_0) \implies \forall q \in \mathbb{R}^+, \quad \mathbb{P}(S_T(h_D^*, f_0, \beta_0) > q) \underset{T \to +\infty}{\longrightarrow} 1.$$

The proof of this result is given in Appendix B.2.

(ii) Correlation with the structural error Another interesting property of the MPI is to be the function of z_{jt} which maximizes the correlation with the structural error.

Proposition 3.4 (Correlation between the MPI and the structural error)
Let $\mathcal{H}$ be the set of measurable functions of z_{jt}, we have under $\overline{H}_a$:

[6] This last expression corresponds to the exact formulation of the MPI under homoskedasticity.
[7] The consistency of the test also holds when we keep the conditional variance term.

$$\forall \alpha \in \mathbb{R}^*, \quad \alpha\ \mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}] \in \arg\max_{h \in \mathcal{H}} \ |corr(\xi_{jt}(f_0, \beta_0), h(z_{jt}))| \,.$$

The proof is given in Appendix B.2. Intuitively, the MPI $h_D^*(z_{jt})$ is designed to fully capture the exogenous variation contained in the correction term $\Delta_{0,a}^{\xi_{jt}}$ implied by the misspecification, which yields the result above.

3.3 Connection with the optimal instruments

The MPI maximizes the power of the moment-based test for $\overline{H}_0 : (f, \beta) = (f_0, \beta_0)$. In contrast, the optimal instruments minimize the asymptotic variance-covariance of the GMM estimator when the parameter of interest is identified by conditional moment restrictions. These two problems are seemingly unrelated. However, we show that the MPI devoted to testing the specification of the model at the true parameter against any fixed local alternative can be rewritten as a linear combination of the optimal instruments. Consequently, one can reinterpret the optimal instruments as a local approximation of the MPI devoted to testing the model at the true parameter. This connection between the MPI and the optimal instruments helps us understand why the feasible approximations of the MPI we construct in section 4 improve the performance of the BLP estimator in our Monte Carlo simulations when the distribution of RCs is flexible. In this subsection, we first derive the optimal instruments. Then, we exhibit the relation between the optimal instruments and the MPI.

The estimation of the model works as follows. The researcher assumes that f belongs to a parametric family $\mathcal{F}_0 = \{f_0(\cdot|\tilde{\lambda}) : \tilde{\lambda} \in \Lambda_0\}$ and wants to estimate the true parameter $\theta_0 = (\beta_0', \lambda_0')'$ under this parametric restriction. In the estimation context that we study here, θ_0 refers to the true parameter. For now, let us assume that the model is correctly specified: $f \in \mathcal{F}_0$ and we shorten the notations by removing the dependence of the structural error in $f_0(\cdot|\tilde{\lambda})$, which becomes implicit in this context. Namely, $\xi_{jt}(f_0(\cdot|\tilde{\lambda}), \tilde{\beta})$ becomes $\xi_{jt}(\tilde{\theta})$. We further assume that θ_0 is point identified by the following moment restriction: $\mathbb{E}[\xi_{jt}(\theta_0)|z_{jt}] = 0$ $a.s.$.[8] The researcher must choose the set of instruments $h_E(z_{jt})$ (or equivalently, the unconditional moments) to include in the GMM objective function:

[8]The identification conditions in the parametric case are less stringent than the conditions for the non-parametric identification in Assumption A.

$$\hat{\theta} = \underset{\tilde{\theta}}{\text{Argmin}}\ TJ \left(\frac{1}{TJ} \sum_{j,t} \hat{\xi}_{jt}(\tilde{\theta}) h_E(z_{jt}) \right)' \hat{W} \left(\frac{1}{TJ} \sum_{j,t} \hat{\xi}_{jt}(\tilde{\theta}) h_E(z_{jt}) \right).$$

Optimal instruments in the BLP demand model Traditionally, the instruments $h_E(z_{jt})$ are chosen to minimize the asymptotic variance-covariance of the estimator $\hat{\theta}$. The instruments that reach this objective are called the optimal instruments. The resulting estimator is said to be efficient in the sense that its asymptotic variance cannot be reduced by using additional moment conditions. There is a large body of literature on the derivation of optimal instruments in econometric models (Amemiya (1974), Chamberlain (1987), Newey (1990, 2004)). The BLP estimator $\hat{\theta}$ is a non-linear GMM estimator and classical results in Chamberlain (1987) and Amemiya (1974) show that the optimal instruments in this case write:

$$h_E^*(z_{jt}) = \mathbb{E}[\xi_{jt}(\theta_0)^2 | z_{jt}]^{-1} \mathbb{E}\left[\frac{\partial \xi_{jt}(\theta_0)}{\partial \tilde{\theta}} \middle| z_{jt} \right],$$

The corresponding efficiency bound (obtained by setting $h_E = h_E^*$) writes:

$$V^* = \mathbb{E}\left[\mathbb{E}\left[\frac{\partial \xi_{jt}(\theta_0)}{\partial \tilde{\theta}} \middle| z_{jt} \right] \mathbb{E}\left[\frac{\partial \xi_{jt}(\theta_0)}{\partial \tilde{\theta}} \middle| z_{jt} \right]' \mathbb{E}[\xi_{jt}(\theta_0)^2 | z_{jt}]^{-1} \right]^{-1}.$$

For the sake of exhaustivity, we show this result in Appendix B.2.1. As for the MPI, the formulation of the optimal instruments above is obtained under the assumption of conditional independence of demand shocks ξ_{jt} in the same market: $k \neq j$, $\xi_{jt} \perp\!\!\!\perp \xi_{kt} | z_t$. In Appendix B.2.1, we derive the expression for the optimal instruments under weaker assumptions on the demand shock.[9] Consistent with what we did in the case of the MPI, we drop the conditional variance term $\mathbb{E}[\xi_{jt}(\theta_0)^2 | z_{jt}]$.

Connection between the MPI and the optimal instruments Let θ_0 the true parameter. Under the parametric assumption $f \in \mathcal{F}_0$, the simple hypothesis $\overline{H}_0 : (f, \beta) = (f_0, \beta_0)$ we studied previously becomes $\overline{H}_0 : \theta = \theta_0$. It is straightforward to show that, in the parametric case, the associated MPI against a fixed alternative

[9] We allow for unrestricted forms of correlation between demand shocks within a given market.

$\overline{H}_a : \theta = \theta_a$ writes: $h_D^*(z_{jt}) = \mathbb{E}\left[\Delta^{\xi_{jt}}_{\theta_0,\theta_a}|z_{jt}\right]$ with $\Delta^{\xi_{jt}}_{\theta_0,\theta_a} = \xi_{jt}(\theta_0) - \xi_{jt}(\theta_a)$. By taking a Taylor expansion of $\xi_{jt}(\theta_a)$ around θ_0, we obtain the following:

$$\Delta^{\xi_{jt}}_{\theta_0,\theta_a} = \frac{\partial \xi_{jt}(\theta_0)}{\partial \tilde{\theta}}(\theta_0 - \theta_a) + o(||\theta_0 - \theta_a||_2)\,.$$

We see that when θ_a is in a neighborhood of θ_0, the MPI, $h_D^*(z_{jt})$, against this fixed alternative is a linear combination of the optimal instruments $h_E^*(z_{jt})$:

$$h_D^*(z_{jt}) = \mathbb{E}\left[\Delta^{\xi_{jt}}_{\theta_0,\theta_a}|z_{jt}\right] \approx \underbrace{\mathbb{E}\left[\frac{\partial \xi_{jt}(\theta_0)}{\partial \tilde{\theta}}\middle| z_{jt}\right]'}_{h_E^*(z_{jt})}(\theta_0 - \theta_a).$$

It follows that classical optimal instruments can be interpreted as an approximation of the MPI devoted to testing $H_0 : \theta = \theta_0$ against any fixed local alternative.[10] Moreover, let us note that the connection between the MPI and the optimal instruments holds if we keep the conditional variance term in both cases.

4 A feasible most powerful instrument

The MPI is the most powerful instrument to reject $\overline{H}_0 : (f, \beta) = (f_0, \beta_0)$ against a fixed alternative $\overline{H}_a : (f, \beta) = (f_a, \beta_a)$. Its derivation requires the knowledge of the alternative while in practice the econometrician typically wants to remain agnostic about the alternative. Moreover, the MPI is defined as a conditional expectation of a non-linear function with respect to a large dimension vector z_{jt}, and thus, even if the alternative $\overline{H}_a$ is known, the MPI can be difficult to compute. In this section, we remain in the same configuration, where the econometrician wants to test $\overline{H}_0 : (f, \beta) = (f_0, \beta_0)$ against a fixed alternative $\overline{H}_a : (f, \beta) = (f_a, \beta_a)$. However now, we assume that this alternative is unknown to the econometrician. We provide two feasible approximations of the MPI, which do not depend on $\overline{H}_a$, and that, unlike the MPI, can be computed in practice. To do so, we show that the MPI can be approximated by a linear combination of known functions of z_{jt}. We call these interval instruments in reference to the way these functions are derived. Our feasible MPI is simply the vector of the interval

[10] This interpretation of the optimal instruments only holds when the model is well specified i.e. $f \in \mathcal{F}_0$, and thus, in general, the optimal instruments shouldn't be used to test the specification of the model.

instruments. The cost to incur for feasibility is that the properties we established for the MPI do not carry over to the feasible MPI. Nevertheless, our Monte Carlo simulations in section 6 show that the interval instruments perform very well in practice.

By construction, in the BLP demand model, the correction term writes:

$$\begin{aligned} \Delta_{0,a}^{\xi_{jt}} &= x_{1jt}'(\beta_a - \beta_0) + \rho_j^{-1}(s_t, x_{2t}, f_0) - \rho_j^{-1}(s_t, x_{2t}, f_a) \\ &= x_{1jt}'(\beta_a - \beta_0) + \Delta_j(s_t, x_{2t}, f_0, f_a). \end{aligned} \tag{4.6}$$

The previous equation shows that the correction term is the sum of a linear part, which is standard, and a non-linear part which is specific to the BLP demand model.

Linear part The linear part of the MPI writes: $\mathbb{E}[x_{1jt}|z_{jt}]'(\beta_a - \beta_0) = \mathbb{E}[x_{1jt}|z_{jt}]'\gamma$. Thus, for its linear part, the MPI is a linear combination of the conditional expectation of x_{1jt} with respect to the exogenous variables with unknown weights. If one is interested in specifically testing that $\beta = \beta_0$, informative instruments simply consist of the variables in $\mathbb{E}[x_{1jt}|z_{jt}]$.

Non-linear part The non-linear part, $\Delta_j(s_t, x_{2t}, f_0, f_a)$, is the part which is implied by the misspecification on the distribution of RCs and for which we need to recover a feasible approximation. Equation (4.6) indicates that the non-linear part is the difference between the inverse demand functions generated by f_0 and f_a. We now go one step further and derive two analytical approximations of $\Delta_j(s_t, x_{2t}, f_0, f_a)$ which we then use as building blocks to construct our feasible approximations of the MPIs. The first approximation is based on a local expansion around f_0. The second approximation is based on an identity that is valid everywhere. The first approximation is more precise locally whereas the second one is more robust to large deviations from f_0.

4.1 Local approximation

First, we consider a local approximation of $\Delta_j(s_t, x_{2t}, f_0, f_a)$. This approximation corresponds to the first order term in the expansion of $\Delta(s_t, x_{2t}, f_0, f_a)$ "around f_0", which is recovered by exploiting the properties of the inverse demand function, which is both $\mathcal{C}^\infty$ and bijective in s_t.

Proposition 4.1
A first order expansion of $\Delta(s_t, x_{2t}, f_0, f_a)$ *around* f_0 *writes:*

$$\Delta(s_t, x_{2t}, f_0, f_a) = \left(\frac{\partial \rho(\delta_t^0, x_{2t}, f_0)}{\partial \delta}\right)^{-1} \int_{\mathbb{R}^{K_2}} \left[\frac{\exp(\delta_t^0 + x_{2t}v)}{1 + \sum_{k=1}^J \exp\left\{\delta_{kt}^0 + x'_{2kt}v\right\}} - \rho(\delta_t^0, x_{2t}, f_0)\right] f_a(v) + \mathcal{R}_0,$$

with $\delta_t^0 = \rho^{-1}(s_t, x_{2t}, f_0)$ *and* $\mathcal{R}_0 = o\left(\int_{\mathbb{R}^{K_2}} |f_a(v) - f_0(v)| dv\right)$.

The proof is in Appendix B.3.1. We first observe that for any density f_0, we can construct artificial market shares s_t^0 such that $\rho^{-1}(s_t, x_{2t}, f_a) = \rho^{-1}(s_t^0, x_{2t}, f_0)$. Then, we recover the final result by taking a Taylor expansion of $\rho^{-1}(s_t^0, x_{2t}, f_0)$ around s_t and showing that the remainder is bounded.[11] This approximation is local by design: it works best when f_a is a local deviation from f_0, even if it can be used more generally. To make this expression useful in practice, we must still overcome two difficulties. The distribution f_a is unknown to the econometrician. In addition, some variables such as δ_{jt}^0 are endogenous. However, notice that the previous expression may be particularly useful if the econometrician is interested in testing $\overline{H}_0$ against a fixed and known alternative as we did in the previous section.

Discretizing the integral To solve for the fact f_a is unknown to the econometrician, we replace the integral in which f_a appears by a finite Riemann approximation. Namely,

$$\int_{\mathbb{R}} \frac{\exp\left\{x'_{2jt}v\right\}}{1 + \sum_{k=1}^J \exp\left\{\delta_{kt}^0 + x'_{2kt}v\right\}} f_a(v) dv \approx \sum_{l=1}^L \omega_l(f_a) \frac{\exp(x'_{2jt}v_l)}{1 + \sum_{k=1}^J \exp(\delta_{kt}^0 + x'_{2kt}v_l)},$$

with $\{v_l\}_{l=1,\ldots,L}$ the points chosen in the domain of definition of f_a, and $\{\omega_l(f_a)\}_{l=1,\ldots,L}$ the associated weights.[12] We provide more details on how to choose the points in Appendix C.4. It is important to observe that in the Riemann approximation, only the weights depend on the alternative f_a. This approximation can also be interpreted as approaching a continuous distribution with a discrete one, where each point in $\{v_l\}_{l=1,\ldots,L}$ represents a specific consumer type with an associated probability $w_l(f_a)$. The non-linear part of the MPI can thus be approximated as follows:

[11] The expansion is taken around s_t because s_t^0 depends on f_a and is thus unknown to the researcher.
[12] In the usual Riemann sum, the weights correspond to density evaluated at point v_l : $f_a(v_l)$ times the width of the interval around v_l.

$$\mathbb{E}[\Delta_j(s_t, x_{2t}, f_0, f_a)|z_{jt}] \approx \sum_{l=1}^{L} \omega_l(f_a)\ \mathbb{E}[\pi_{j,l}(s_t, x_t)|z_{jt}],$$

$$\text{with } \pi_{j,l}(s_t, x_t) = \left(\frac{\partial \rho(\delta_t^0, x_{2t}, f_0)}{\partial \delta}\right)^{-1} \left[\frac{\exp(\delta_t^0 + x_{2t}v_l)}{1 + \sum_{k=1}^{J} \exp\{\delta_{kt}^0 + x'_{2kt}v_l\}} - \rho(\delta_t^0, x_{2t}, f_0)\right]_j .$$

Approximating the conditional expectation Ideally, we would like to estimate the conditional expectation of $\pi_{j,l}(s_t, x_t)$ with respect to z_{jt}. The endogenous variables are $\{\delta_{jt}^0\}_{j=1,...,J}$, and the potential endogenous variables in $\{x_{2jt}\}_{j=1,...,J}$, which often include prices. In practice, computing the conditional expectation is challenging because the dimension of z_{jt} can be very large and the functions $\pi_{j,l}(\cdot)$ are highly non-linear and non-separable in the endogenous variables. This makes it unappealing to use standard non-parametric estimation methods.[13] In the same spirit as Reynaert and Verboven (2014), we first project the endogenous variables on the space spanned by a relevant subset of z_{jt}. We mark the projected endogenous variables with a hat and we plug them into our functions $\pi_{j,l}(\cdot)$. Namely, we have the following approximation for every interval instrument l:

$$\mathbb{E}[\pi_{j,l}(s_t, x_t)|z_{jt}] \approx \hat{\pi}_{j,l}(z_{jt}) = \left(\frac{\partial \rho(\hat{\delta}_t^0, \hat{x}_{2t}, f_0)}{\partial \delta}\right)^{-1} \left[\frac{\exp(\hat{\delta}_t^0 + \hat{x}_{2t}v_l)}{1 + \sum_{k=1}^{J} \exp\left\{\hat{\delta}_{kt}^0 + \hat{x}'_{2kt}v_l\right\}} - \rho(\hat{\delta}_t^0, \hat{x}_{2t}, f_0)\right]_j .$$

We show in Appendix C.2 that this strategy yields an estimator of the conditional expectation that converges faster to a first order approximation of the conditional expectation.

Test procedure From what precedes, the MPI (for its non-linear part) can be approximated as follows: $h_D^*(z_{jt}) \approx \sum_{l=1}^{L} \omega_l(f_a)\ \hat{\pi}_{j,l}(z_{jt})$. As we don't know the weights $\omega_l(f_a)$, we propose to take the vector $\hat{\pi}_j(z_{jt}) = (\hat{\pi}_{j,1}(z_{jt}), ..., \hat{\pi}_{j,L}(z_{jt}))'$ as our testing instruments. We call them interval instruments in reference to the way we divide the support into several intervals to construct this approximation. Following

[13] For instance, a Sieve nonparametric estimator of the conditional mean. The dimension of z_{jt} makes this approach of little relevance in practice.

the test procedure presented in section 3.1, we perform a moment based test for $\overline{H}_0 : \mathbb{E}\left[\hat{\pi}_j(z_{jt})\xi_{jt}(f_0,\beta_0)\right] = 0$. Under the same assumptions as in Proposition 3.1 and setting $h_D(z_{jt}) = \hat{\pi}_j(z_{jt})$, we have the following:

$$\text{Under } H_0 : S_T(h_D, f_0, \beta_0) \xrightarrow[T\to+\infty]{d} \chi^2_L.$$

This approach has the advantage of being feasible since we can construct the vector of interval instruments $\hat{\pi}_j(z_{jt})$, while remaining completely agnostic about f_a. The price to pay is that we lose the optimality properties of the MPI. We further discuss the properties of the feasible MPI in Appendix C.7. Moreover, the infeasible MPI, $h^*_D(z_{jt})$, is of dimension one and its test statistic is distributed as χ^2_L asymptotically. In contrast, the feasible MPI is of dimension L and its asymptotic distribution is a χ^2_L. This increase in the number of degrees of freedom may lead to some loss of power. An alternative approach would consist in letting the researcher choose the weights $\{\hat{\omega}_l\}_{l=1,\ldots,L}$ and recover an instrument of dimension one. However, for this approach to work well and retain good power properties, the econometrician must choose the weights so that they approximately match the real weights $\{w_l(f_a)\}_{l=1,\ldots,L}$. This requires a good prior knowledge of the cumulative distribution function of the alternative distribution f_a. Nevertheless, our Monte Carlo simulations in section 6 show that the feasible MPIs that we propose perform very well in practice.

4.2 Global approximation

Second, we consider a global approximation that is based on an identity which is valid everywhere and not only when f is close to f_a. Simple algebraic operations (see Appendix B.3.2) allow us to derive the following expression for $\Delta_j(s_t, x_{2t}, f_0, f_a)$. Let $\delta^0_{jt} = \rho^{-1}_j(s_t, x_{2t}, f_0)$ and $\delta^a_{jt} = \rho^{-1}_j(s_t, x_{2t}, f_a)$. We have:

$$\Delta_j(s_t, x_{2t}, f_0, f_a) = \log\left(\frac{\int_{\mathbb{R}^{K_2}} \frac{\exp(x'_{2jt}v)}{1+\sum_{k=1}^{J}\exp\left\{\delta^a_{kt}+x'_{2kt}v\right\}} f_a(v)dv}{\int_{\mathbb{R}^{K_2}} \frac{\exp(x'_{2jt}v)}{1+\sum_{k=1}^{J}\exp\left\{\delta^0_{jt}+x'_{2kt}v\right\}} f_0(v)dv}\right).$$

As for the local approximation, we cannot directly exploit this formula as some quantities such as f_a and δ^a_{jt} are unknown and some variables such as δ^0_{jt} are endogenous. To remedy these two difficulties, we apply the same methods as previously described: we discretize the integral, and we project the endogenous variables onto the space spanned by a relevant subset of z_{jt}. To solve for the fact that the mean utility δ^a_{jt} under the alternative is unknown, we replace it with the mean utility under the null δ^0_{jt}. This should not alter the approximation too much given that δ^a_{jt} only enters the expression at the denominator within a sum, which averages out the differences between δ^a_{jt} and δ^0_{jt} across products. In the end, we are able to provide the following approximation for the non-linear part of the MPI:

$$\mathbb{E}[\Delta_j(s_t, x_{2t}, f_0, f_a)|z_{jt}] \approx \log\left(\sum_{l=1}^{L} \bar{\omega}_l(f_a)\ \hat{\bar{\pi}}_{j,l}(z_{jt})\right) \text{ with } \hat{\bar{\pi}}_{j,l}(z_{jt}) = \frac{\frac{\exp(x'_{2jt}v_l)}{1+\sum_{k=1}^{J}\exp\left\{\hat{\delta}^0_{kt}+x'_{2kt}v_l\right\}}}{\int_{\mathbb{R}^{K_2}} \frac{\exp(x'_{2jt}v)}{1+\sum_{k=1}^{J}\exp\left\{\hat{\delta}^0_{jt}+x'_{2kt}v\right\}} f_0(v)dv},$$

where $\{\bar{\omega}_l(f_a)\}_{l=1,...,L}$ correspond to the unknown weights and the $\hat{\bar{\pi}}_{j,l}(z_{jt})$ are set of global interval instruments. The MPI can thus be approximated by the logarithm of a weighted sum of known functions of z_{jt}. As we did previously, we use $\hat{\bar{\pi}}_j(z_{jt}) = (\hat{\bar{\pi}}_{j,1}(z_{jt}), ..., \hat{\bar{\pi}}_{j,L}(z_{jt}))'$ as instruments to test $\overline{H}_0$. All the weights are positive and sum to one, which entails that the non-linear part of the correction term is an increasing function of our instruments. This approximation is said to be global because contrary to the second approximation we study, it does not require f_0 to be close to f_a. Nevertheless, if f_a is close to f_0, then the fraction κ inside the logarithm is close to 1 and the well-known approximation $\log(\kappa) \approx \kappa - 1$ allows us to directly rewrite the MPI as a linear combination of our instruments.

Overall, the feasible MPIs that we derive in this section allows us to approximate the most powerful instrument against a fixed alternative while remaining agnostic about this alternative.

4.3 Feasible MPIs for estimation

In the estimation framework, the researcher stipulates that f belongs to a parametric family $\mathcal{F}_0 = \{f_0(\cdot|\tilde{\lambda}) : \tilde{\lambda} \in \Lambda_0\}$ and wants to estimate the true parameter $\theta_0 = (\beta'_0, \lambda'_0)'$ under this parametric restriction. From the connection between the MPI and the local instruments that we present in section 3.3, we can infer that good estimation instruments $h_E(z_{jt})$ ought to approximate the MPI devoted to testing $H_0 : \theta = \theta_0$ against any local alternative. If we have an initial estimator of θ_0, we can directly use the interval instruments presented previously to approximate the MPI devoted to testing $H_0 : \theta = \theta_0$ against an unknown alternative. The fact that the feasible MPIs do not depend on the alternative is key for estimation. Moreover, the transformation of the MPI into a vector of instruments of dimension $L \geq |\lambda_0|$ is necessary for estimation as the number of instruments must be greater than the dimension of the parameter to estimate.[14] In Appendix C.5, we propose a version of the interval instruments that does not require a first step estimate of θ_0 and that can be computed directly from the logit specification.

5 Composite hypothesis

In the traditional estimation procedure, which encompasses almost all the applications of the BLP model, the econometrician must make a parametric assumption on the distribution of random coefficients to estimate the model. Formally, the econometrician assumes f belongs to a parametric family $\mathcal{F}_0 = \{f_0(\cdot|\tilde{\lambda}) : \tilde{\lambda} \in \Lambda_0\}$, where $\tilde{\lambda}$ is a parameter that must be estimated. In applied work, researchers typically assume that f is normally distributed. This parametric choice is rarely grounded in economic theory and, if too restrictive, is likely to impose arbitrary restrictions on some key counterfactual quantities such as the pass-through. In this section, we develop a formal specification test for $H_0 : f \in \mathcal{F}_0$. In comparison to the test in section 3.1, we must now estimate the parameters of the distribution $\theta_0 = (\beta_0', \lambda_0')'$ in a first step, which generates parameter uncertainty. Moreover, we propose a rigorous treatment of the numerical approximations involved in the derivation of the structural error $\xi_{jt}(\tilde{\theta})$. We organize this section as follows. First, we define the pseudo-true value associated with a given specification and the first stage estimator. Second, we define our test procedure and its implementation in practice. Finally, we study the asymptotic properties of our test.

5.1 Pseudo-true value and first stage estimator

To estimate the BLP model, researchers must make three choices. They must choose the parametric family $\mathcal{F}_0$, the instruments $h_E(z_{jt})$ to estimate the model, and a weighting matrix W, which weights the different moments included in the objective function. Given these three choices, we can define the BLP pseudo-true value $\theta(\mathcal{F}_0, h_E, W) \equiv$

[14]The linear parameter β_0 has its own instruments, which are simply the variables in x_{1jt}.

$\theta_0 = (\beta_0', \lambda_0')'$ as follows:[15]

$$\theta(\mathcal{F}_0, h_E, W) \in \underset{\tilde{\theta}}{\text{Argmin}}\ \mathbb{E}\left[\xi_{jt}(f_0(\cdot|\tilde{\lambda}), \tilde{\beta}) h_E(z_{jt})\right]' W \mathbb{E}\left[h_E(z_{jt}) \xi_{jt}(f_0(\cdot|\tilde{\lambda}), \tilde{\beta})\right].$$

If the model is well-specified ($f \in \mathcal{F}_0$) and the pseudo-true value is unique, then the pseudo-true value is the true value: $\theta_0 = \theta$. Under misspecification, θ_0 is a parameter whose value depends on $(\mathcal{F}_0, h_E, W)$. For exposition purposes, we omit this dependence in the subsequent analysis. Moreover, here we remain general and do not impose that W must be equal to the usual optimal weighting matrix. It is often the case in practice, that the researchers choose the identity matrix or regularize the weighting matrix.

First stage estimator $\hat{\theta}$ The first stage estimator is an empirical counterpart of the BLP pseudo-true value defined previously. The minimization is done with respect to sample analogs. Additionally, we know that there is no closed form expressions for the structural error $\xi_{jt}(f_0(.|\tilde{\lambda}), \tilde{\beta})$, and thus, we must use a feasible counterpart $\hat{\xi}_{jt}(f_0(.|\tilde{\lambda}), \tilde{\beta})$ instead.

$$\hat{\theta}(\mathcal{F}_0, h_E, \hat{W}) \equiv \hat{\theta} = \underset{\tilde{\theta}}{\text{Argmin}} \left(\sum_{j,t} \hat{\xi}_{jt}(f_0(.|\tilde{\lambda}), \tilde{\beta}) h_E(z_{jt})\right)' \hat{W} \left(\sum_{j,t} \hat{\xi}_{jt}(f_0(.|\tilde{\lambda}), \tilde{\beta}) h_E(z_{jt})\right). \tag{5.7}$$

The construction of the feasible structural error $\hat{\xi}_{jt}(f_0(.|\tilde{\lambda}), \tilde{\beta})$ requires the following 3 numerical approximations:

1. The econometrician does not observe a continuum of consumers as in the theoretical model but only empirical averages $\hat{s}_{jt}$ over the n_t individuals in market t.
$$\hat{s}_{jt} = \frac{1}{n_t} \sum_{i=1}^{n_t} y_{ijt}, \tag{5.8}$$
where $y_{ijt} \in \{0; 1\}$ are i.i.d. choices over the $i = 1, \dots, n_t$.

2. There is no closed form for $\rho_j(., x_{2t}, f_0(\cdot|\tilde{\lambda}))$, the integral has to be computed

[15] Our definition of a pseudo-true value is closely related to the approach in White (1982) in the context of maximum likelihood. In his case, the pseudo true value minimizes the Kullback-Leibler distance between the assumed likelihood and the true likelihood, whereas in our case, the pseudo-true value minimizes a weighted sum of population moments.

through numerical integration. A prominent example is Monte Carlo integration:

$$\hat{\rho}_j(\delta, x_{2t}, f_0(|\tilde{\lambda})) = \frac{1}{R}\sum_{r=1}^{R}\frac{\exp(\delta_j + x_{2jt}v_r)}{1+\sum_{k=1}^{J_t}\exp(\delta_k + x'_{2kt}v_r)}, \tag{5.9}$$

with v_r iid draws from $f_0(\cdot|\tilde{\lambda})$.

3. There is no analytical way to recover the inverse of the demand functions $\rho^{-1}(s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))$. The most popular way to derive the inverse demand is by solving the following contraction:

$$C : (\cdot, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) : \delta \mapsto \delta + \log(s_t) - \log(\rho(\delta, x_{2t}, f_0(\cdot|\tilde{\lambda}))).$$

   This solution has given rise to the popular nested fixed point GMM procedure.[16]

In Section 5.3, we explicitly state the assumptions that allow us to neglect these approximations asymptotically.

5.2 Test procedure

Under Assumption A, and assuming $h_E(z_{jt})$ and W are such that the pseudo-true value θ_0 is unique, the following equivalence holds:

$$\begin{aligned} H_0 : f \in \mathcal{F}_0 &\iff \overline{H}_0 : (f, \beta) = (f_0(\cdot|\lambda_0), \beta_0) \\ &\iff \mathbb{E}[\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0)|z_{jt}] = 0 \quad a.s.. \end{aligned}$$

The pseudo true value reduces the dimensionality of the problem by allowing us to move from a composite hypothesis $H_0 : f \in \mathcal{F}_0$ to the simple hypothesis $\overline{H}_0 : (f, \beta) = (f_0(\cdot|\lambda_0), \beta_0)$ studied previously. As we did in section 2, we propose a moment-based test of H_0.[17] Under H_0, for every set of testing instruments $h_D(z_{jt})$, the following moment conditions must hold:

$$H_0 : f \in \mathcal{F}_0 \iff \overline{H}_0 : (f, \beta) = (f_0(\cdot|\lambda_0), \beta_0) \implies \overline{H}'_0 : \mathbb{E}\left[\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0)h_D(z_{jt})\right] = 0.$$

[16] Another solution that has gained traction in the literature is the MPEC procedure (Dubé et al. (2012)) that replaces the BLP inversion at each step of the minimization by imposing equilibrium constraints on the minimization program.

[17] Other testing approaches could have been considered. First, one could use the previous equivalence to directly test H_0 via an integrated conditional moment test. We do not follow this route for at least two reasons. First, this test will contain no information on the nature of the misspecification (it could be completely unrelated to the distribution of RC). Second, in practice the dimension of z_{jt} is often very large, which substantially reduces the power of this kind of test. Another testing approach would

We now develop a procedure to test $\overline{H}'_0$. In comparison to the test in section 3.1, we must now account for the fact that the pseudo-true value needs to be estimated to derive the test statistic, which generates parameter uncertainty. Moreover, we propose a rigorous treatment of the numerical approximations involved in the derivation of the structural error.

Test statistic For any choice of testing instruments $h_D(z_{jt})$, our objective is to test $\overline{H}'_0 : \quad \mathbb{E}[\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0) h_D(z_{jt})] = 0$ where $\theta_0 = (\beta'_0, \lambda'_0)'$ is the pseudo-true value associated with the parametric family $\mathcal{F}_0$.[18] In order to test $\overline{H}_0$, we consider the following Wald test statistic:

$$S_T(h_D, \mathcal{F}_0, \hat{\theta}) = TJ \left(\frac{1}{TJ} \sum_{j,t} \widehat{\xi}_{jt}(f_0(\cdot|\hat{\lambda}), \hat{\beta}) h_D(z_{jt}) \right)' \hat{\Sigma} \left(\frac{1}{TJ} \sum_{j,t} \widehat{\xi}_{jt}(f_0(\cdot|\hat{\lambda}), \hat{\beta}) h_D(z_{jt}) \right).$$

where $\hat{\Sigma}$ is a weighting matrix chosen by the econometrician and $\hat{\theta} = (\hat{\beta}, \hat{\lambda})$ is a consistent estimator of θ_0. The number of markets T is the dimension that we let grow to infinity to the asymptotic properties of our test. We motivate this choice in Appendix C.3. Under some regularity conditions that we make explicit in the following section, the asymptotic distribution of the test statistic under $\overline{H}'_0$ is as follows:

$$S_T(h_D, \mathcal{F}_0, \hat{\theta}) \xrightarrow{d} Z' \Sigma Z, \tag{5.10}$$

$$\text{with } \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{j=1}^{J} \widehat{\xi}_{jt}(f_0(\cdot|\hat{\lambda}), \hat{\beta}) h_D(z_{jt}) \xrightarrow{d} Z \sim \mathcal{N}(0, \tilde{\Omega}_0). \tag{5.11}$$

Σ is the probability limit of $\hat{\Sigma}$. We make $\tilde{\Omega}_0$ explicit in the next subsection (in particular, the derivation of $\tilde{\Omega}_0$ takes into account parameter uncertainty). Given that $\hat{\Sigma}$ is chosen by the econometrician and it is possible to derive a consistent estimator of $\tilde{\Omega}_0$,

have entailed testing $H_0 : f \in \mathcal{F}_0$ against a larger class of densities that encompasses $\mathcal{F}_0$. For instance, if $\mathcal{F}_0$ is the family of normal distributions, encompassing families are mixtures of Gaussians with a larger number of components. We do not follow this route for two reasons. First, it is not desirable to restrict the alternative to a class of distributions that encompass the null as the econometrician does not know a priori the misspecification. Second, estimating the BLP model with a more flexible parametrization is challenging. An advantage of our test procedure is that it doesn't require estimating the model with a more flexible parametrization.

[18] Remember that under an alternative specification, the pseudo true value also depends on the estimation instruments $h_E(z_{jt})$ and the weighting matrix.

the econometrician can always simulate the asymptotic distribution of the test statistic. In some polar cases, which we present hereafter, the asymptotic distribution of our test statistic is pivotal chi-square distribution that does not require to be simulated.

Two polar cases For the sake of exposition, let us now describe two polar cases where the asymptotic distributions are pivotal chi-square distributions, which do not require to be simulated. Denote by $|\cdot|_0$ the counting norm.

1. **Sargan-Hansen J test:** If the set of estimation instruments and the set of testing instruments are the same ($h_E = h_D$), if $\hat{W}$ is the 2-step GMM optimal weighting matrix and if $\hat{\Sigma} = \hat{W}^{-1}$, then our test boils down to the usual Sargan-Hansen J test and we have under $\overline{H}_0'$:

$$S_T(h_D, \mathcal{F}_0, \hat{\theta}) \xrightarrow{d} \chi^2_{|h_E|_0 - |\theta|_0}.$$

2. **Non-redundant h_D and h_E:** if $\tilde{\Omega}_0$ has full rank and if the econometrician sets $\hat{\Sigma} = \hat{\tilde{\Omega}}_0^{-1}$, then our test statistic has the following asymptotic distribution under $\overline{H}_0'$:[19]

$$S_T(h_D, \mathcal{F}_0, \hat{\theta}) \xrightarrow{d} \chi^2_{|h_D|_0}.$$

Choice of the testing instruments As previously indicated, the power properties of our test hinge critically on the choice of testing instruments $h_D(z_{jt})$. We established that the MPI and its feasible counterparts, the interval instruments, feature attractive properties in testing $\overline{H}_0 : (f, \beta) = (f_0(\cdot|\lambda_0), \beta_0)$ against any fixed alternative. Thus, it is natural to use these instruments for the specification test above. In particular, we show that the consistency of the test with the MPI carries over to the general specification test above in Appendix B.5.

5.3 Asymptotic validity

We now study the asymptotic properties of our test when the number of markets T goes to infinity. To establish the asymptotic validity and consistency of our test, we

[19] If Ω_0 is singular, one can always use directly the asymptotic distribution in 5.10 or apply the singularity-robust procedure proposed in Andrews and Guggenberger (2019).

exploit classical results on the asymptotic normality of the non-linear GMM estimator (Hansen (1982), Newey (1990)) as well on the large-T asymptotics of the BLP estimator (Freyberger (2015)). The main challenge here is to control the magnitude of the approximations that intervene in the derivation of the structural error so that they can be neglected asymptotically. Contrary to Freyberger (2015), we do not assume the convergence of any moments ex-ante and we allow for the approximation error between demand and observed market shares to be non-zero.

Assumption B
(i) $(s_t, x_t, z_t)_{t=1}^T$ are i.i.d. across markets and are consistent with the probability model defined by equations (2.1), (2.2) and (2.3) evaluated at (f, β);
(ii) Strong Exogeneity: $\mathbb{E}[\xi_{jt}(f, \beta)|z_{jt}] = 0$ a.s.;
(iii) Finite moment conditions: x_{2t} has bounded support and x_{1t} has finite 4th moments.

In B(i), we assume that the data are i.i.d. across markets, an assumption which we could relax slightly (technically, only certain moments need to be identical across markets), and that the data are generated by the BLP demand model at a given pair (f, β). In B(ii), we assume exogeneity of our instrumental variables. Let us stress that to show the asymptotic validity of our specification test, we do not require (f, β) to be non-parametrically identified, as we just need parametric identification under H_0. In particular, we do not need all the assumptions in A. B(iii) is a necessary condition to recover the asymptotic normality of the BLP estimator. x_{1t} having finite 4th moments is standard. x_{2t} having bounded support has two purposes. First, it implies that the structural error has a finite 4th moment, Compiani (2018) makes the same assumption on price for this purpose. Second, it ensures that the mapping used in the nested fixed point algorithm is a proper smooth contraction, which allows us to prove that the NFP algorithm converges (without truncating the contraction mapping as in Berry (1994) and Berry et al. (1995)) and control for the NFP approximation bias.

Assumption C
$\mathcal{F}_0$ is such that :
(i) λ_0 belongs to the interior of Λ_0 with Λ_0 compact;
(ii) $\tilde{\lambda} \mapsto \rho(\delta, x_{2t}, f_0(\cdot|\tilde{\lambda}))$ is well defined and continuously differentiable on Λ_0.

In C(i), we assume that, for any given DGP, the associated pseudo-true-value λ_0 associated with the family $\mathcal{F}_0$ lies in a compact space Λ_0. This condition is standard in establishing the consistency and asymptotic normality of M-estimators. Second, in C(ii), we impose that the demand function and its derivative with respect to λ should both be well defined and continuous.

Next, we impose conditions on the instruments that are used for estimation $h_E(z_{jt})$ and for testing $h_D(z_{jt})$ and on the BLP estimator itself.

Assumption D
For a given $\mathcal{F}_0$ that satisfies Assumption C and for some weighting matrix W and Σ, the following conditions must hold:
(i) Finite moments for instruments: $h_E(z_{jt})$ and $h_D(z_{jt})$ are not perfectly colinear and have finite 4th moments;
(ii) Global identification of θ_0: $\exists!\theta_0$ such that $\forall\tilde{\theta} \neq \theta_0$:

$$\mathbb{E}\left[\sum_j \xi_{jt}(f_0(\cdot|\tilde{\lambda}),\tilde{\beta})h_E(z_{jt})'\right] W\mathbb{E}\left[\sum_j h_E(z_{jt})\xi_{jt}(f_0(\cdot|\tilde{\lambda}),\tilde{\beta})\right] > \mathbb{E}\left[\sum_j \xi_{jt}(f_0(\cdot|\lambda_0),\beta_0)h_E(z_{jt})'\right] W\mathbb{E}\left[\sum_j h_E(z_{jt})\xi_{jt}(f_0(\cdot|\lambda_0),\beta_0)\right];$$

(iii) Local identification: $\Gamma(\mathcal{F}_0,\theta_0,h_E) = \mathbb{E}\left[\sum_j h_E(z_{jt})\frac{\partial\xi_{jt}(f_0(\cdot|\lambda_0),\beta_0)}{\partial\theta'}\right]$ and $\Gamma(\mathcal{F}_0,\theta_0,h_D)$ have full column rank;
(iv) W and Σ are symmetric positive definite and $\hat{W} \xrightarrow{P} W$, $\hat{\Sigma} \xrightarrow{P} \Sigma$;
(v) $\hat{\theta}$ minimizes objective function (5.7) and satisfies the FOC of the minimization problem:

$$\left(\sum_{j,t}\frac{\partial\hat{\xi}_{jt}(f(\cdot|\hat{\lambda}),\hat{\beta})}{\partial\theta}h_E(z_{jt})\right)'\hat{W}\left(\sum_{j,t}\hat{\xi}_{jt}(f(\cdot|\hat{\lambda}),\hat{\beta})h_E(z_{jt})\right) = 0.$$

Assumption D restricts the class of instruments which can be used for estimation and for testing. More specifically, D(i) and D(iii) are common regularity conditions necessary to establish asymptotic results whereas D(ii) is an identification condition which ensures that the pseudo true value θ_0 is uniquely defined, which is critical to show the consistency of the BLP estimator. Finally, Assumptions D(iv) and D(v) impose regularity conditions on the weighting matrix as well as on the BLP estimator itself.

The next assumptions ensure that the numerical approximations involved in the derivation of the structural error do not interfere with the asymptotic theory.

Assumption E
(i) Let n_t be the number of individuals in market t, $(n_t)_{t=1}^T$ is i.i.d. and independent from all other variables. First, it must be that $\forall t \quad \sqrt{T}\mathbb{E}(n_t^{-1/2}) \underset{T\to+\infty}{\to} 0$. Second, observed market share $\hat{s}_t$ in market t must write:

$$\hat{s}_{jt} = \frac{1}{n_t}\sum_{i=1}^{n_t} y_{ijt},$$

with $(y_{ijt})_{i=1}^{n_t}$ i.i.d. draws generated by the BLP demand model at a given pair (f, β) conditional on (x_t, ξ_t).
(ii) Let R be the number of simulations, then the simulated demand for product j writes:

$$\hat{\rho}_{jt}(\delta, x_{2t}, f_0(\cdot|\tilde{\lambda})) = \frac{1}{R}\sum_r \frac{\exp(\delta_j + x'_{2jt}v_r)}{1 + \sum_k \exp(\delta_k + x'_{2kt}v_r)},$$

where $v_r \overset{iid}{\sim} f_0(\cdot|\tilde{\lambda})$, and $\frac{T}{R} \underset{T\to+\infty}{\to} 0$.
(iii) Let H be the stopping time for the contraction (which depends on T) and ϵ the fixed Lipschitz constant of the contraction mapping used to invert the demand function, then it must be that $\sqrt{T}\epsilon^H \underset{T\to+\infty}{\to} 0$.

A sufficient condition for E(i) to hold is that the minimum number of individuals observed in any market is of higher order than the total number of markets. This condition can be checked in practice.[20] Assumptions E(ii) and E(iii) can also be checked in practice and are more manageable because R and H are chosen by the researcher and can always be increased so that these assumptions hold.

Given our assumptions, we derive the asymptotic distribution of our test statistic under the null, and show that the test is consistent.

Theorem 5.1 *Let $\hat{\theta} = \hat{\theta}(\mathcal{F}_0, \hat{W}, h_E)$ be the BLP estimator associated with distributional assumption $\mathcal{F}_0$, weighting matrix $\hat{W}$, estimating instruments h_E. Under assumptions B-E,*

[20]Note that by making stronger assumptions on the higher moments and the support of the observed characteristics, it is possible to find milder conditions on the number of individuals relative to the number of markets.

- *Under* $\overline{H}'_0$: $\mathbb{E}\left[\xi_{jt}(f_0(\cdot|\lambda_0),\beta_0)h_D(z_{jt})\right]=0$,

$$S_T(h_D,\mathcal{F}_0,\hat{\theta}) \underset{T\to+\infty}{\overset{d}{\to}} Z'\Sigma Z, \qquad\qquad Z\sim\mathcal{N}(0,\tilde{\Omega}_0),$$

$$\textit{where}\quad \tilde{\Omega}_0=\begin{pmatrix} I_{|h_D|0} & G\end{pmatrix}\begin{pmatrix}\Omega(\mathcal{F}_0,h_D) & \Omega(\mathcal{F}_0,h_D,h_E)\\ \Omega(\mathcal{F}_0,h_D,h_E)' & \Omega(\mathcal{F}_0,h_E)\end{pmatrix}\begin{pmatrix} I_{|h_D|0}\\ G'\end{pmatrix},$$

$$\Omega(\mathcal{F}_0,h_D,h_E)=cov\left(\sum_j \xi_{jt}(f(.|\lambda_0),\beta_0)h_D(z_{jt}),\sum_j \xi_{jt}(f(.|\lambda_0),\beta_0)h_E(z_{jt})\right),$$

$$G=-\Gamma(\mathcal{F}_0,\theta_0,h_D)\left[\Gamma(\mathcal{F}_0,\theta_0,h_E)'W\Gamma(\mathcal{F}_0,\theta_0,h_E)\right]^{-1}\Gamma(\mathcal{F}_0,\theta_0,h_E)'W.$$

- *Under* H'_a : $\mathbb{E}\left[h_D(z_{jt})\xi_{jt}(f_0(.|\lambda_0),\beta_0)\right]\neq 0$,

$$\forall q\in\mathbb{R}^+,\quad \mathbb{P}(S_T(h_D,\mathcal{F}_0,\hat{\theta})>q)\underset{T\to+\infty}{\to} 1.$$

The proof of Theorem 5.1 is in Appendix B.4 and comprises three main steps. First, we show that under the assumptions in E, the numerical approximation becomes asymptotically negligible. Second, we show the consistency and asymptotic normality of the BLP estimator. Finally, we derive the asymptotic distribution of the test statistic, taking into account parameter uncertainty (θ_0 is estimated and not observed). The apparent complexity of the asymptotic variance-covariance matrix Ω_0 is a direct consequence of parameter uncertainty.

6 Monte Carlo experiments

In this section, we conduct three distinct sets of Monte Carlo experiments. First, we implement a simple simulation exercise to assess the effects of incorrectly specifying the distribution of random coefficients on quantities of interest such as price elasticities or cross-price elasticities, which are known to play a key role in shaping the counterfactuals. In a second set of Monte Carlo experiments, we study the finite sample performances of the specification test developed in section 5 with different sets of testing instruments. We first examine the size of our test in finite sample. Then, we investigate the power properties of our test under alternative specifications (with alternatives including Gaussian mixtures, gamma distributions and local alternatives). We show that our test with the interval instruments significantly outperforms the traditional J-test with the usual instruments. Finally, in the last Monte Carlo exercise, we study the performance of the interval instruments to estimate the parameters of the model by means of comparison with the commonly used instruments in the literature.

6.1 Simulation design

For the sake of exposition, we will keep the same simulation design for all the simulation experiments. The simulation design closely follows the simulation design used in Dubé et al. (2012), Reynaert and Verboven (2014). The market includes $J = 12$ products, which are characterized by 3 exogenous product attributes x_a, x_b and x_c that follow a joint normal distribution. The price p is endogenous and depends on the observed and unobserved characteristics and on some cost shifters c_1 and c_2. Consumer heterogeneity is present only in x_c, and the random coefficient v_i associated with x_c follows various distributions depending on the simulation exercise. The sample size T varies between 50, 100 and 200 markets. We can summarize the DGP as follows:

$$u_{ijt} = 2 + x_{ajt} + 1.5x_{bjt} - 2p_{jt} + x_{cjt}v_i + \xi_{jt} + \varepsilon_{ijt} \qquad \xi_{jt} \sim \mathcal{N}(0,1),\ \varepsilon_{ijt} \sim EV1,$$

$$\text{and} \begin{bmatrix} x_{a,j} \\ x_{b,j} \\ x_{c,j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -0.8 & 0.3 \\ -0.8 & 1 & 0.3 \\ 0.3 & 0.3 & 1 \end{bmatrix} \right),$$

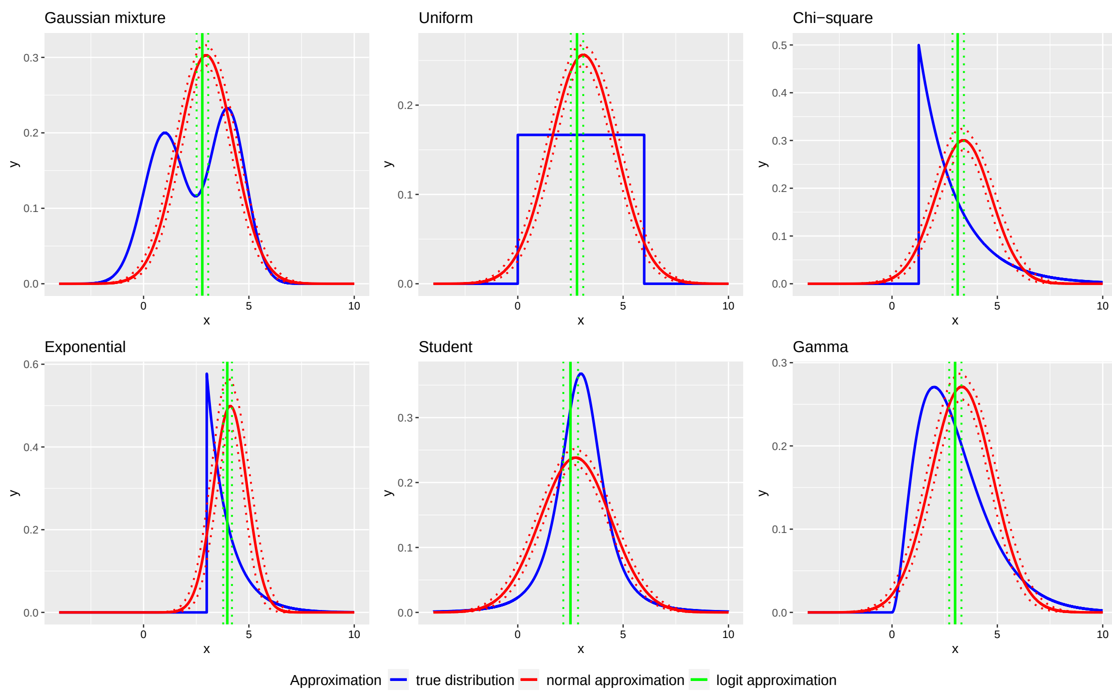
$$p_{jt} = 1 + \xi_{jt} + u_{jt} + \sum_{k=a}^{c} x_{kjt} + c_{1jt} + c_{2jt} \quad \text{with } u_{j,t} \sim U[-4,-2],\ c_{1jt} \sim U[2,4]\ and\ c_{2jt} \sim U[3,5].$$

Market shares are generated by integrating over 20,000 consumers. This allows us to essentially remove the approximation error between the observed and theoretical market shares.

6.2 Counterfactuals under an alternative distribution

We now present a simple exercise to illustrate how the misspecification of random coefficients can affect the estimation of quantities of interest such as price elasticities and cross-price elasticities. To do so, we simulate data using the simulation design introduced above and we take various distributions for the random coefficient v_i (respectively: Gaussian mixture, Uniform, Chi-square, Exponential, Student, Gamma). We ensure that all the distributions have the same mean and variance (3 and 3, respectively). For each distribution, we simulate $T = 100$ markets of data and we estimate the model either assuming no heterogeneity (simple logit) or assuming that v_i is normally distributed. We replicate the same exercise 500 times for each distribution. This allows us to recover the mean estimate for the parameters as well as to construct 95% "confidence bins" (by trimming the observations below the 2.5% quantile and above the 97.5% quantile). We plot the true densities and their estimated counterparts under the normal and logit assumptions in Figure 1. We observe that the estimated logit parameters and the estimated means of the normal distributions always coincide and are close to 3 for all the distributions. However, there is some variation between the different specifications. For instance, the estimated means are larger with the exponential distribution. The estimated variances also vary from one specification to the other. The estimated variances for the exponential distribution are smaller, while they are larger for the student distribution.

Figure 1: True densities and estimated densities under normal and logit specifications



In a second stage, we simulate $N = 5,000$ draws from the true distributions as well as from the estimated logit and normal approximations to compute the demand, the price-elasticity and the cross-price elasticity for the product j^* with the highest value for x_c.[21] The cross-price elasticity is arbitrarily taken for product $j = 1$ with respect to p_{j^*}. We derive the quantities of interest for 100 equally spaced values of p_{j^*} ranging in $]0, 10[$. We plot the elasticities in Figure 2 and cross-price elasticities in Figure 3 generated by the true distribution as well as those generated by the logit and normal approximations, respectively. We proceed similarly with the demand functions. We see in Figure 9 in Appendix).

One can observe that, as expected, the logit specification poorly replicates the substitution patterns. In particular, it consistently overstates the magnitude of the elasticities and cross-elasticities with respect to the true ones. The absence of consumer heterogeneity on characteristic c implies that consumers can "renounce' more easily to product j^* when its price increases. By introducing some heterogeneity, the normal approximation somewhat attenuates this issue. However, significant discrepancies in

[21]The expressions for both price-elasticities and the cross-price elasticities are in Appendix D.1.

the shape of elasticities and cross-price elasticities remain. As most counterfactual analyzes rely on the substitution patterns generated by the model, these differences will inevitably create significant biases.

Figure 2: Price elasticities

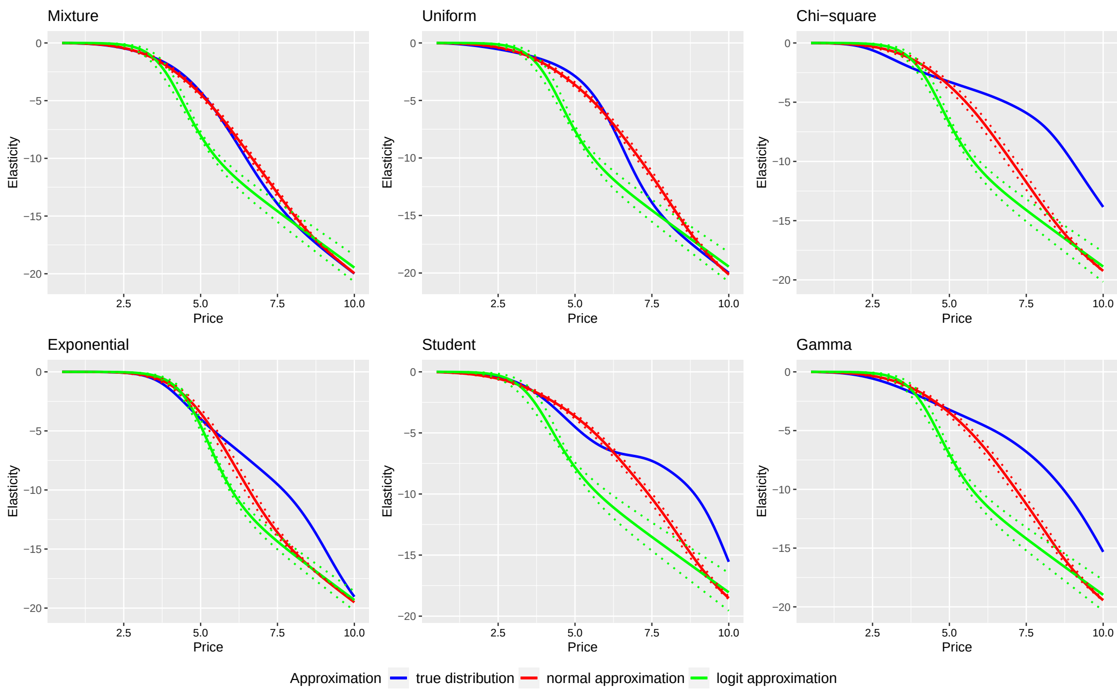
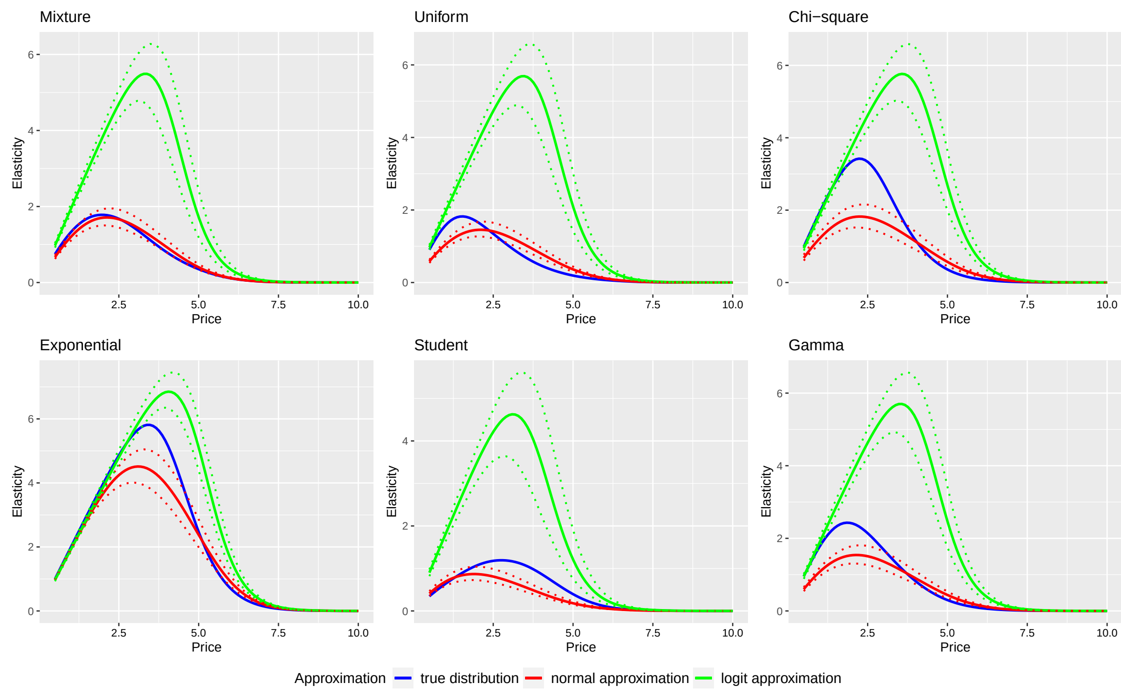


Figure 3: Cross-price elasticities



6.3 Finite sample performance of the specification test

We now study the empirical size and power of our test under different sample sizes and for different sets of testing and estimating instruments. Once again, the data are generated according to the simulation design exhibited previously for various distributions of v_i. The assumption made throughout the simulations is $H_0 : f \in \mathcal{F}_0$, where $\mathcal{F}_0$ is the family of normal distributions. In other words, we always assume that the random coefficient is normally distributed and we test this hypothesis. We set the nominal size to 5%. We study the finite sample performances of the specification test that we presented in section 5 using different sets of estimation and testing instruments. For estimation, we take the instruments commonly adopted by practitioners: the differentiation instruments of Gandhi and Houde (2019) and the "optimal" instruments of Reynaert and Verboven (2014). Both of these sets are approximations of the classical optimal instruments. Second, we compare the performance of the test when performing the standard Sargan-Hansen J test (i.e. when we use the same instruments for testing and estimation) and when we use the global and local approximations of the MPI that we constructed in sections 4.2 and 4.1. We denote the latter tests as I

Local and I Global respectively. The BLP estimator is computed following the NFP GMM procedure described in section 5.1. For the optimization, only an analytic Jacobian is provided. We ensure that the number of tested restrictions is of the same magnitude across the different sets of instruments. More details on the exact sets of instruments and on the estimation procedure for this specific set of simulations are given in Appendix D.2.

6.3.1 Empirical size

The size is the probability of rejecting the null hypothesis when the null is true, so we compute the empirical size by counting and averaging the number of times we reject the null for nominal size 5% over the $1,000$ simulations when the random coefficient v_i is normally distributed. Below in Table 1, we report the empirical sizes of the test with the different sets of instruments described above for the different sample sizes $T \in \{50, 100, 200\}$ and for different distributions of the RC such that $v_i \sim f \in \mathcal{F}_0$.

Table 1: Empirical size for nominal size 5% (1000 replications)

Number of markets	T=50						T=100						T=200					
Estimation instruments	"Differentiation"			"Optimal'			"Differentiation"			"Optimal'			"Differentiation"			"Optimal'		
Test type	J	I Global	I local	J	I Global	I local	J	I Global	I local	J	I Global	I local	J	I Global	I local	J	I Global	I local
$v_i \sim \mathcal{N}(-1, 0.5^2)$	0.294	0.083	0.091	0.145	0.078	0.063	0.138	0.078	0.058	0.094	0.084	0.047	0.08	0.052	0.053	0.064	0.05	0.04
$v_i \sim \mathcal{N}(0, 0.75^2)$	0.293	0.084	0.085	0.148	0.081	0.071	0.137	0.061	0.06	0.1	0.059	0.05	0.074	0.053	0.045	0.062	0.048	0.036
$v_i \sim \mathcal{N}(1, 1^2)$	0.287	0.084	0.083	0.142	0.084	0.073	0.142	0.055	0.054	0.098	0.053	0.047	0.079	0.042	0.03	0.058	0.035	0.025
$v_i \sim \mathcal{N}(2, 2^2)$	0.288	0.087	0.077	0.145	0.071	0.072	0.138	0.069	0.051	0.099	0.053	0.056	0.077	0.044	0.041	0.069	0.037	0.044
$v_i \sim \mathcal{N}(3, 3^2)$	0.287	0.089	0.071	0.137	0.075	0.066	0.145	0.074	0.06	0.098	0.06	0.061	0.076	0.044	0.037	0.061	0.046	0.046

We observe that with a moderate sample size ($T = 50, J = 12$), all the tests are over-sized. This is within expectations and due to the approximations inherent to the estimation of the BLP models as described in section 5 and the relatively large number of instruments used for estimation and testing purposes.[22] However, we notice that the Sargan-Hansen J tests are much more over-sized than the tests with the interval instruments: the rejection rate is above 25% for the Sargan-Hansen J test with

[22]The number of over-identifying restrictions lies between 6 and 8. The Sargan-Hansen J tests are known to suffer from size distortions as the number of instruments increases.

differentiation instruments vs around 8% for the I test. Increasing the sample size improves the tests' empirical levels and shifts them towards the nominal level, which is a good indication of the validity of our test. Even with a relatively large number of markets ($T = 200$), the Sargan-Hansen J tests remain slightly oversized (rejection rate is still slightly above 5%). On the other hand, for the test with interval instruments, the empirical size appears to match the nominal level for all but two configurations, where it seems to be slightly undersized.

6.3.2 Empirical power

Power is the probability of rejecting the null hypothesis under an alternative. We compute the empirical power by counting and averaging the number of times we reject the null for the test of nominal size 5% over the 1000 simulations when the distribution of random coefficients is misspecified. The simulation setup remains the same as previously with the only modification being that the true distribution of v_i is now either a mixture of normals or a Gamma. We report the power against the different alternatives in the subsequent tables. The main takeaway from our results is that the test with the interval instruments as testing instruments (I global and I local) largely outperforms the traditional Sargan-Hansen J-test against all the alternative distributions considered in our simulations.

Power against Gaussian mixture alternatives We simulate data with the random coefficients distributed according to the Gaussian mixtures described below. We plot the true distributions in Figure 4. We report the results in Table 2. We observe that the test with the interval instruments has great power against all the mixtures tested. The rejection rates go to 1 very quickly in comparison to the Sargan-Hansen J tests.

$$v = Dv_1 + (1-D)v_2, \quad \mathbb{P}(D=1) = p, \quad \mathbb{P}(D=0) = 1-p,$$
$$v_1 \sim \mathcal{N}\left(-\sqrt{\frac{3p}{1-p}} + 2, 1\right) \quad v_2 \sim \mathcal{N}\left(\sqrt{\frac{3(1-p)}{p}} + 2, 1\right),$$

with $p \in \{0.1; 0.2; 0.3; 0.4; 0.5\}$.

Figure 4: Densities, Gaussian mixture alternatives

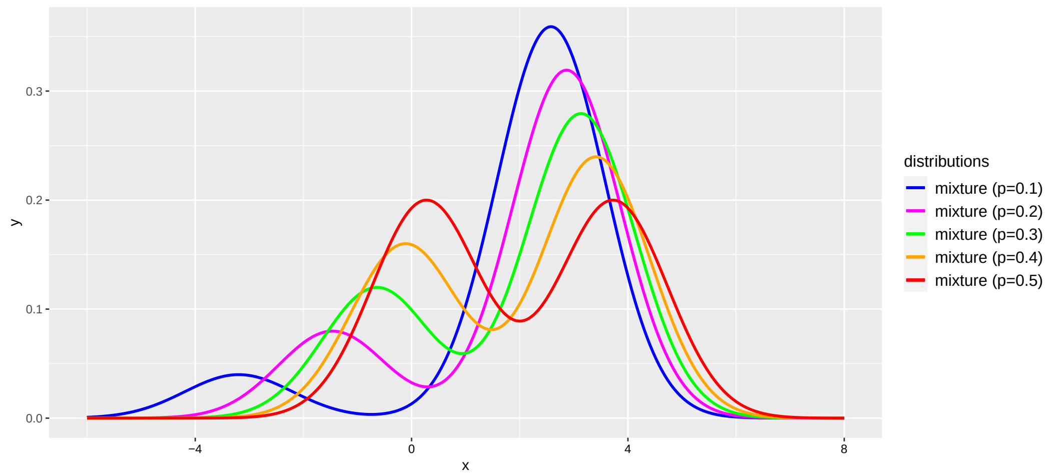


Table 2: Empirical power, Gaussian mixture alternatives (1000 replications)

Number of markets	T=50						T=100						T=200					
Estimation instruments	"Differentiation"			"Optimal'			"Differentiation"			"Optimal'			"Differentiation"			"Optimal'		
Test type	J	I Global	I Local	J	I Global	I Local	J	I Global	I Local	J	I Global	I Local	J	I Global	I Local	J	I Global	I Local
Mixture 1	0.533	0.991	0.987	0.719	0.989	0.989	0.604	1	1	0.967	1	1	0.829	1	1	1	1	1
Mixture 2	0.626	0.996	0.998	0.613	0.997	0.998	0.723	1	1	0.905	1	1	0.933	1	1	1	1	1
Mixture 3	0.629	0.992	0.995	0.43	0.996	0.997	0.741	1	1	0.7	1	1	0.941	1	1	0.977	1	1
Mixture 4	0.601	0.983	0.982	0.275	0.981	0.981	0.713	1	0.999	0.368	1	1	0.921	1	1	0.672	1	1
Mixture 5	0.56	0.907	0.904	0.157	0.9	0.906	0.635	0.993	0.995	0.124	0.995	0.996	0.855	1	1	0.146	1	1

Power against Gamma alternatives We simulate data with the random coefficients distributed according to the Gamma distribution described below. We plot the true distributions in Figure 5. We report the results in table 3. We observe that the test with interval instruments has great power against all the Gamma distributions tested except for the first one, which we can see on the plot has a distribution that is relatively close to a normal distribution. Even for the first Gamma distribution, it still outperforms the traditional sets of instruments. For all the other Gamma distributions, the rejection rates go to 1 very quickly in comparison to the Sargan-Hansen J-tests. This confirms the superiority of the interval instruments in detecting misspecification in the distribution of random coefficients. In Appendix D.2, we also study the power properties of our test against local alternatives.

$$v \sim \Gamma(2, k) \quad with\ k \in \{0.25, 0.5, 0.75, 1, 1.5\}$$

Figure 5: Densities, Gamma alternatives

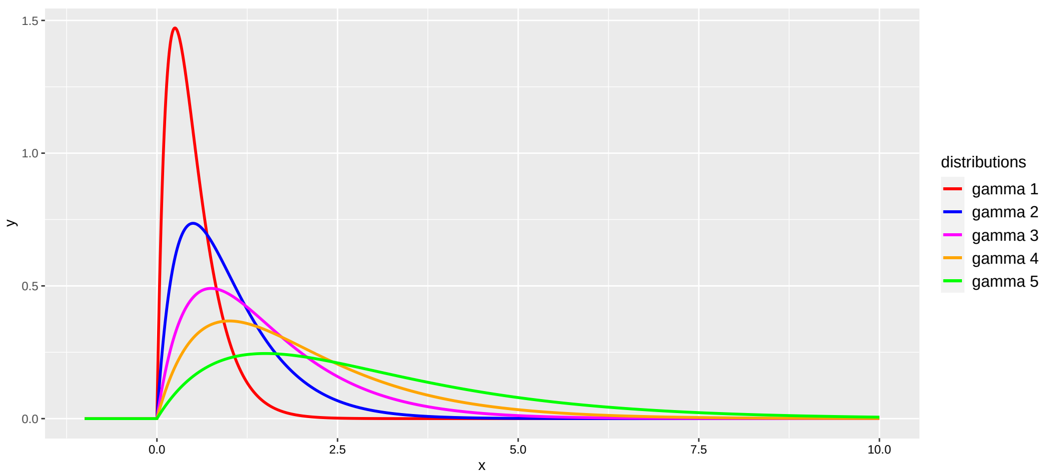


Table 3: Empirical power, Gamma alternatives (1000 replications)

Number of markets	T=50						T=100						T=200					
Estimation instruments	"Differentiation"			"Optimal'			"Differentiation"			"Optimal'			"Differentiation"			"Optimal'		
Test type	J	I Global	I Local	J	I Global	I Local	J	I Global	I Local	J	I Global	I Local	J	I Global	I Local	J	I Global	I Local
Gamma 1	0.293	0.106	0.093	0.142	0.082	0.074	0.154	0.083	0.073	0.094	0.092	0.08	0.118	0.155	0.139	0.066	0.156	0.138
Gamma 2	0.516	0.747	0.752	0.14	0.781	0.77	0.562	0.983	0.978	0.095	0.982	0.98	0.492	1	1	0.08	1	1
Gamma 3	0.607	0.96	0.962	0.157	0.963	0.969	0.693	0.998	1	0.156	1	1	0.922	1	1	0.161	1	1
Gamma 4	0.622	0.97	0.99	0.207	0.962	0.995	0.748	0.999	1	0.263	1	1	0.933	1	1	0.412	1	1
Gamma 5	0.687	0.991	0.999	0.371	0.988	0.999	0.812	1	1	0.585	1	1	0.976	1	1	0.865	1	1

6.4 Finite sample performance of interval instruments for estimation

In our last simulation exercise, we evaluate the performance of our interval instruments in estimating the parameters associated with the RC when the distribution of random coefficients is flexibly parametrized. To do so, we simulate data with a distribution of random coefficients following a mixture of Gaussians and we estimate the parameters of this mixture. We provide a method to estimate the parameters when the distribution of the RC is a mixture in section C.6 of the Appendix. In particular, we provide a new parametrization of the model, which yields substantial practical gains and may be of interest to researchers independent of the rest of the paper. The simulation design remains the same as previously. We assume that the random coefficient v_i is distributed according to the following mixture: $v_i \sim D_i\ \mathcal{N}(-2, 0.5) + (1 - D_i)\ \mathcal{N}(4, 0.5)$ with $\mathbb{P}(D_i = 1) = 0.25$. Thus, there are 5 parameters associated with the distribution of RC: the means and variances of each component of the mixture and the mixing probability. Our objective is to compare the performance of the global and local interval instruments with the instruments commonly used by practitioners: the differentiation instruments from Gandhi and Houde (2019) and the "optimal instruments" from Reynaert and Verboven (2014). In Table 4, we report the empirical biases and the square root of the MSE for the estimators of the non-linear parameters for each set of instruments and for the different sample sizes. In Appendix D.3, we report the same information for the linear parameters (see Tables 14, 15, and 17) as well as the distribution of the empirical distribution of the non-linear estimates. Table 4 allows us to directly compare the performances of the three sets of instruments in estimating the non-linear parameters. We first observe that for all the sets of instruments, the empirical biases and $\sqrt{MSE}$ of the estimators decrease when the sample size increases, which is reassuring. Furthermore, it appears clearly that the differentiation instruments perform worse than the "optimal instruments" and the interval instruments. The empirical $\sqrt{MSE}$ of the estimators with the differentiation instruments is up to 12 times larger than with the interval instruments and up to 6 times larger than with the "optimal instruments". We reach the same conclusions when we study empirical biases. The interval instruments appear to perform better than the "optimal instruments" even if the difference is less significant than with the differentiation instruments. For the sake of conciseness, we do not report the results obtained with a mixture of 3 components but the observations we make with two components are even more exacerbated. In Appendix D.3, as a means of comparison, we perform the same exercise when the distribution of random coefficients is a simple Gaussian and here, we do not observe any significant differences between the different sets of instruments, which confirms that the interval instruments make a difference when the distribution of RCs is flexible.

Table 4: Estimation non-linear parameters of the mixture (1000 replications)

	Instruments	Differentiation					"Optimal"					Interval Global					Interval Local				
	Parameter	β_{3L}	σ_{3L}	β_{3H}	σ_{3H}	p_L	β_{3L}	σ_{3L}	β_{3H}	σ_{3H}	p_L	β_{3L}	σ_{3L}	β_{3H}	σ_{3H}	p_L	β_{3L}	σ_{3L}	β_{3H}	σ_{3H}	p_L
Sample size	true	-2	0.5	4	0.5	0.25	-2	0.5	4	0.5	0.25	-2	0.5	4	0.5	0.25	-2	0.5	4	0.5	0.25
T=50, J=12	bias	0.214	0.184	-0.022	-0.045	0.027	0.076	0.059	0.026	-0.111	0.01	0.017	0	-0.045	0.004	0.005	-0.006	-0.005	-0.039	-0.001	0.003
	$\sqrt{MSE}$	0.633	0.734	0.281	0.35	0.075	0.361	0.483	0.212	0.281	0.036	0.277	0.391	0.227	0.259	0.024	0.251	0.34	0.214	0.244	0.019
T=50, J=20	bias	0.189	0.347	0.022	-0.081	0.025	0.074	0.11	0.028	-0.089	0.01	0.013	0.042	-0.018	-0.003	0.004	0.019	0.033	-0.023	0.01	0.003
	$\sqrt{MSE}$	0.566	0.887	0.184	0.291	0.059	0.328	0.563	0.163	0.228	0.033	0.248	0.415	0.166	0.22	0.021	0.228	0.38	0.15	0.184	0.018
T=100, J=12	bias	0.233	0.226	0.02	-0.066	0.027	0.054	0.037	0.019	-0.066	0.007	0.004	-0.012	-0.027	0.005	0.002	0	0	-0.028	0.007	0.001
	$\sqrt{MSE}$	0.592	0.703	0.256	0.305	0.072	0.279	0.4	0.154	0.211	0.028	0.167	0.282	0.157	0.201	0.013	0.127	0.225	0.143	0.164	0.005
T=100, J=20	bias	0.198	0.423	0.047	-0.101	0.025	0.074	0.107	0.033	-0.074	0.01	-0.009	-0.005	-0.008	-0.009	0.001	-0.003	0.004	-0.01	0.004	0.001
	$\sqrt{MSE}$	0.552	0.89	0.164	0.27	0.055	0.311	0.52	0.129	0.194	0.034	0.115	0.264	0.115	0.169	0.005	0.104	0.226	0.103	0.125	0.004
T=200, J=12	bias	0.184	0.167	0.011	-0.049	0.019	0.026	0.011	0.021	-0.061	0.004	-0.006	-0.027	-0.015	-0.001	0.001	0.002	-0.007	-0.016	0.006	0.001
	$\sqrt{MSE}$	0.466	0.601	0.176	0.262	0.053	0.184	0.313	0.113	0.172	0.018	0.088	0.219	0.108	0.164	0.003	0.091	0.174	0.099	0.123	0.003

7 Empirical application

The objective of the empirical exercise is twofold. First, we want to verify how well our instruments perform at estimating a flexible distribution of RCs using a real data set. Second, we want to study how the shape of the distribution of RCs can modify key counterfactual quantities such as the price elasticities or the pass-through, and check whether the results we obtain are consistent with the findings in Miravete et al. (2022). To do so, we estimate demand for cars using data on new car registrations in Germany from 2012 to 2018.[23] There are many reasons to focus on the car market. First, cars are highly differentiated products, which makes the BLP framework particularly adapted to this market. As a result, the BLP demand model has been widely applied to study the car industry (e.g., Berry et al. (1995), Grigolon et al. (2018), Petrin (2002)) and one can easily compare our results with previous results obtained in the literature under different specifications. Second, there are many policy-relevant questions related to this market. In particular, the role of road transport in air pollution is significant and many countries have implemented tax policies to reduce the CO2 emissions generated by car transportation.[24] An important strand of the literature has investigated

[23]The dataset was kindly provided to us by Kevin Remmy https://kevinremmy.com/research/.

[24]In 2017, road transport was responsible of approximately 19% of total greenhouse has emissions in EU-28 Retrieved from https://www.eea.europa.eu/data-and-maps/indicators/

the performance of these different taxation schemes (Alberini and Horvath (2021), Allcott and Wozny (2014), D'Haultfœuille, Givord, and Boutin (2014), Durrmeyer (2022), Durrmeyer and Samano (2018), Gillingham and Houde (2021), Grigolon et al. (2018), Huse and Koptyug (2022), Kunert (2018)). Other policy-relevant questions include the impact of import tariffs (Miravete et al. (2018)) and the determinants of market power (Berry et al. (1995), Grieco, Murry, and Yurukoglu (2022)). To answer these questions, the researcher must often estimate the demand for cars. The credibility of the implied analysis depends critically on how well the model can reproduce the underlying substitution patterns and the shape of the demand curve. To this end, it is essential to have a demand system that is sufficiently flexible, and particularly so with respect to the random coefficient on price. In this section, we use our instruments to estimate a Gaussian mixture as the random coefficient associated with price. Moreover, we use our test to assess how moving from the usual Gaussian RC to the Gaussian mixture decreases the degree of misspecification. Finally, we compare the counterfactual quantities under a Gaussian mixture and the traditional specifications (Gaussian RC and logit). In line with the findings in Miravete et al. (2022), our results indicate that the Gaussian mixture yields higher pass-through rates and curvatures.

7.1 Data

The data set includes state-level new car registrations, publicly available by the German Federal Motor Transport Authority (KBA) from 2012 to 2018. This gives us 112 markets defined by state-year pairs. Data on car characteristics and price are scraped from General German Automobile Club and include horsepower, engine type, size, weight, fuel cost, CO2 emission, number of doors, segment, and body type. The data set is at a granular level where every car is uniquely identified by its manufacturer and its type key code (HSN/TSN) that is defined according to the characteristics of the car. Following the literature, we aggregate products with the same brand, model, engine type, and body combination (e.g. BMW-1 Series-Diesel-Hatchback).[25] Likewise, we follow the literature and define the market size as the number of households in the market. To construct market shares, we simply divide new car registrations of a

`transport-emissions-of-greenhouse-gases/transport-emissions-of-greenhouse-gases-12` on October 21, 2022.

[25] In aggregating the products from the HSN/TSN level, we use the characteristics of the car with the highest sales.

given product by the market size. The data set is complemented by information on demographics such as the number of households or the average income per household at the state-year level and yearly average gas price data from ADAC.[26]

Summary statistics Shares of products sold by engine type are presented in Table 5. We focus our analysis on combustion engine vehicles as in our sample period electric-vehicle cars constitute a small market share (always less than 5% of the sold vehicles) and can be seen as a distinct market. Between diesel and gasoline cars, we observe that the market share for diesel decreases over time, starting from 2016. The timing is in line with the emissions scandal, known as the Dieselgate, which started in September 2015.

Table 5: Shares (%) of new registrations by engine type

	Year						
Fuel Type	2012	2013	2014	2015	2016	2017	2018
Diesel	46.8	46.1	46.3	46.4	43.9	36.2	30.0
Gasoline	52.6	52.9	52.6	52.3	54.4	60.8	66.5
Battery EV	0.1	0.2	0.3	0.4	0.3	0.7	1.1
Hybrid EV	0.5	0.8	0.7	0.6	1	1.4	1.6
Plug-in hybrid EV	0	0	0.1	0.3	0.4	0.9	0.9

Table 6 provides sales-weighted averages for prices and observed characteristics. We observe that the difference in fuel consumption and resulting fuel costs steadily ranks diesel above gasoline. However, the average price of diesel cars sold is higher than gasoline cars. This implies a potential trade-off in terms of the costs of car ownership at the time of purchase. With a fixed mileage in mind, a consumer with high sensitivity to fuel costs might be willing to pay a higher price for a more fuel-efficient car. We also observe that the horsepower and the size of the newly registered cars increase over time.

[26]State level income https://ec.europa.eu/eurostat/web/products-datasets/-/nama_10r_2hhinc

Table 6: Summary Statistics (Sales weighted)

	Year						
	2012	2013	2014	2015	2016	2017	2018
Diesel							
Price/income	0.74	0.72	0.73	0.72	0.71	0.69	0.68
Size (m2)	8.31	8.31	8.32	8.36	8.42	8.48	8.53
Horsepower (kW/100)	1.09	1.07	1.11	1.11	1.14	1.16	1.21
Fuel cost (euros/100km)	7.90	7.18	6.63	5.53	4.94	5.25	5.83
Fuel cons. (Lt./100km)	5.19	4.98	4.89	4.73	4.61	4.61	4.71
CO2 emission (g/km)	136.19	130.50	127.69	123.58	120.42	120.49	123.27
Nb. of products/market	133	138	146	150	151	149	143
Gasoline							
Price/income	0.46	0.46	0.46	0.46	0.46	0.45	0.43
Size (m2)	7.23	7.27	7.28	7.30	7.36	7.46	7.53
Horsepower (kW/100)	0.79	0.78	0.80	0.82	0.85	0.88	0.91
Fuel cost (euros/100km)	9.48	8.61	8.11	7.27	6.69	7.06	7.40
Fuel cons. (Lt./100km)	5.76	5.47	5.40	5.31	5.25	5.34	5.38
CO2 emission (g/km)	135.80	128.18	125.27	122.89	121.22	122.86	123.26
Nb. of products/market	157	171	179	185	186	193	188

Note: Provided statistics are sales weighted averages across products. Total number of markets (State*Year) is 112 .

Inter-market variation Our dataset contains both geographical variation and time variation, as we observe the sales in every state in Germany over the period 2012-2018. States in Germany differ significantly in terms of income per inhabitant, population density and average distance driven.[27] It is fundamental to take this inter-market variation into account in our empirical specification for two reasons. First, our model

[27]For the population density 2019 (inh/km^2): 69 (Mecklenburg-Vorpommern) to 4118 (Berlin) (from Federal Statistical Office of Germany (Destatis)), GDP per capita 2019: 28.9k (Mecklenburg-Vorpommern) to 67k (Hamburg) (retrieved from https://www.ceicdata.com/en/germany/esa-2010-gdp-per-capita-by-region/gdp-per-capita-bremen on 05 November 2022). For average driving distance in 2019: 13079 km (Mecklenburg-Vorpommern) to 9531 (Berlin) retrieved from https://de.statista.com/statistik/daten/studie/644381/umfrage/fahrleistung-privater-pkw-in-deutschland-nach-bundesland/ on 19 September 2022.

postulates that consumers' preferences are the same across markets. However, we observe that the market shares vary from one state to the other even if the choice set remains the same. This feature of the data can only be explained if we let the preferences vary from one market to the other. Second, in section 2.3, we saw that there needs to be sufficient variation in the product characteristics across markets to identify the distribution of RCs. By interacting product characteristics with state demographics, we achieve both objectives: we shift the preferences to a more common representation and we create variation in the product characteristics across markets. To choose which interaction terms to include in the utility function, we first create market specific sales-weighted characteristics for the following variables: price, fuel cost, size, horsepower, height, gasoline dummy, and foreign dummy (equal to one if the manufacturer of the car is not German). Then, we regress these quantities on the demographics of interest: average income, population density, and a time trend. Last, we select the interaction terms that explain the best the variation in sales-weighted characteristics (namely, the variables with a p-value below $1e^{-10}$). The results of these regressions are presented in Table 7. They suggest that income shifts positively the preferences for price, size, and horsepower (i.e. higher income is associated with larger cars, and higher horsepower). In contrast, income shifts negatively the preferences for foreign status, height, and gasoline status.[28] Although weaker, a similar pattern is observed for the effect of population density on car characteristics.

[28] In the main analysis, we use price/income to capture the income effect.

Table 7: Linear regressions of sales-weighted car characteristics on demographic characteristics

	Income(/1000)	Population density (/100)	Time trend
Price($\times$1000)	0.138**	0.069*	0.286*
	(0.013)	(0.011)	(0.059)
Fuel cost (euros/100km)	-0.0069	-0.0036	0.3587**
	(0.0063)	(0.0056)	(0.0296)
Size(m^2)	0.0058**	0.0018*	0.0176*
	(0.00079)	(0.00070)	(0.00371)
Horsepower (KW/100)	0.0028**	0.0012*	0.0129**
	(0.00028)	(0.00025)	(0.00132)
Foreign	−0.0050**	−0.0014*	0.0295**
	(0.00052)	(0.00046)	(0.00246)
Height(m)	−0.00051**	−0.00043**	0.00181*
	(0.000061)	(0.000054)	(0.000286)
Gasoline	−0.0067**	−0.0024*	0.0131*
	(0.00059)	(0.00053)	(0.00280)

Note: * p-value lower than 0.01, ** p-value lower than $1e^{-10}$.

Instruments for the endogeneity of price To instrument for price, we use a combination of variables on the intensity of competition and cost shifters. To measure the intensity of competition, we consider the number of competing products of the same class and engine type in a given market, and the number of competing products of the same engine type in a given market. As for cost shifters, we use three complementary datasets: the mean hourly labor cost, the price of steel (interacted with the weight of the car), and exchange rates between Germany and the country of assembly.

1. Labor cost: we use the mean nominal hourly labor cost per employee in the manufacturing sector of the country of assembly of the models. The data on labor costs come from International Labor Organization Statistics (ILOSTAT).[29]

2. Price of steel: we collect the price of steel futures in January of each year.

3. Exchange rates: we construct the exchange rates between Germany and the country of assembly of each car model using exchange rate data from OECD.[30]

[29] https://www.ilo.org/ilostat-files/Documents/Excel/INDICATOR/LAC_4HRL_ECO_CUR_NB_A_EN.xlsx

[30] https://data.oecd.org/conversion/exchange-rates.htm

7.2 Empirical specification

The indirect utility of consumer i, purchasing product j in market t (defined as a state-year pair) is given by:

$$u_{ijt} = \underbrace{x'_{1jt}\beta + \xi^*_{jt}}_{\delta_{jt}} + x'_{2jt}\alpha_i + \varepsilon_{ijt}.$$

The mean utility $\delta_{jt} = x'_{1jt}\beta + \xi^*_{jt}$ captures homogeneous preferences. The variables in x_{1jt} consist of the product characteristics for which we assume that there is no preference heterogeneity and the interaction terms that explain the best the geographical variation observed in Table 7.[31]

The demand shock on product j is decomposed as follows:

$$\xi^*_{jt} = Brand_j + State_t + Year_t + \xi_{jt},$$

where $Brand_j$ is a brand fixed effect that captures the unobserved quality of the brand of product j, $State_t$ captures state specific demand shocks that are fixed across time and products and $Year_t$ captures year-specific demand shocks. Therefore, $State_t$ and $Year_t$ play a role in explaining the variation in the overall demand for cars (or equivalently, in the share of the outside option).

The variables in x_{2jt} are the product characteristics that display preference heterogeneity and which we augment with a RC. In our specification, we include the price, the size, and the gasoline dummy in x_{2jt}. We estimate the model assuming different specifications for the distribution of RCs. First, we estimate the model without any consumer heterogeneity. Second, we assume that all the RCs are normally distributed. Finally, we consider a Gaussian mixture on price to increase flexibility with respect to the preferences on price. For each different specification, we perform the specification test developed in section 5 to see how the degree of misspecification evolves as we increase flexibility on the distribution of RCs.

[31] The choice of the variables that display preference heterogeneity is based on our understanding of the car market and follows current empirical practices for this specific market. However, we understand the limitations of this approach, and we are working on an iterative procedure to select the variables that display consumer heterogeneity.

7.3 Estimation

Estimation conditional logit (no heterogeneity) First, we estimate the logit model, and we report the results in Table 8.[32] As expected, we find a negative effect of price and fuel cost on the utility. The interaction terms indicate that the utility derived from size, horsepower, foreign status and gasoline all decrease with income. Moreover, we observe that the aversion to fuel cost decreases over time, which is likely an artifact implied by increasing fuel cost over the years. In contrast, the utility derived from horsepower appears to increase with time. However, these time effects are smaller in comparison with the heterogeneity due to income. To facilitate the interpretation of these results, we consider a household with a €47,000 income in 2018. This corresponds to the mean income in 2018. For this household, the implied effect of size on the utility is negative, whereas a positive utility is derived from higher horsepower, the car's brand being German, height, and gasoline engines.

Table 8: Logit estimation

	Baseline		**× Income (/1000)**		**× Pop. density(/100)**		**× Time trend**	
Homogeneous Preferences	$\hat{\beta}$	S.E	$\hat{\beta}$	S.E	$\hat{\beta}$	S.E	$\hat{\beta}$	S.E
Price/income	-2.4	1.3e-01	-	-	-	-	-	-
Fuel Cost	-0.25	8.6e-03	-	-	-	-	0.014	1.7e-03
Size(m^2)	0.15	4.2e-02	-0.0055	8.5e-04	-	-	-	-
Horsepower(KW/100)	2.7	1.8e-01	-0.019	2.4e-03	-	-	-0.081	7e-03
Foreign	0.18	7.1e-02	-0.017	1.4e-03	-	-	-	-
Height(m)	3.5	2.3e-01	-0.0015	4.6e-03	-0.036	4.7e-03	-	-
Gasoline	1.1	6.3e-02	-0.011	1.2e-03	-	-	-	-

Note: Brand, Year and State FE's are included.

Estimation with Gaussian random coefficients We now increase the flexibility in the *traditional* manner, by assuming that the RCs on the price, the size and the gasoline indicator follow a Gaussian distribution. We report the estimates obtained under this new specification in Table 9. The signs for the homogeneous preference

[32] In Appendix E, we provide results for baseline specifications including the simple conditional logit and the nested logit (with and without state and year fixed effects).

parameters in x_{1jt} remain the same and the magnitude of the effects do not change significantly. The sign associated with the mean effect of price remains negative. In contrast, the sign on the mean effects of the size and the gasoline dummy are inverted with respect to the logit specification. This last observation illustrates an important empirical finding: average effects are not invariant to the introduction of preference heterogeneity. In other words, the logit estimates do not necessarily match the means, when we introduce a Gaussian RC. Moreover, the three RCs display high variances and particularly so for the gasoline dummy, which indicate a high level of heterogeneity with respect to these three characteristics.[33]

Table 9: Traditional BLP (Gaussian RC)

	Baseline		**× Income (/1000)**		**× Pop. density(/100)**		**× Time trend**	
Homogeneous Preferences	$\hat{\beta}$	S.E	$\hat{\beta}$	S.E	$\hat{\beta}$	S.E	$\hat{\beta}$	S.E
Price/income	-	-	-	-	-	-	-	-
Fuel Cost	-0.29	5.1e-03	-	-	-	-	0.031	9.2e-04
Size(m^2)	-	-	-0.0053	3.1e-04	-	-	-	-
Horsepower(KW/100)	0.77	1.5e-02	0.0078	6.8e-04	-	-	-0.12	5.6e-03
Foreign	0.21	5.4e-02	-0.019	1.1e-03	-	-	-	-
Height(m)	3.4	1.1e-02	-0.0088	1.2e-03	-0.032	3.6e-04	-	-
Gasoline	-	-	-0.0028	8.6e-04	-	-	-	-
Gaussian RC	$\hat{\beta}$	S.E	$\hat{\sigma}$	S.E				
Price/income	-2.4	2e-02	0.96	5.9e-03	-	-	-	-
Size(m^2)	-0.37	1.5e-02	0.43	3.6e-03	-	-	-	-
Gasoline	-2.3	4.4e-02	4	4.1e-04	-	-	-	-

Note: Brand, Year and State FE's are included.

Estimation with a Gaussian mixture on the price Finally, we increase the flexibility of the model, by replacing the Gaussian RC on the price variable with a Gaussian mixture of 2 components. We focus on the price as the literature shows that the distribution of price sensitivity is absolutely key for many quantities of interest

[33] The estimation is performed using the parametrization proposed in Ketz (2019), which avoids boundary issues at 0 for the variances of the RCs.

in IO, including the price elasticities and the pass-through. We report the estimates obtained under this new specification in Table 10. The results point out the presence of two distinct modes in the distribution of the RC associated with price. The two modes reveal the presence of two groups of consumers: the first one with high price sensitivity (with the mean component at -9.6) and the second one with low price sensitivity (with the mean component at -2.5). Moreover, the distribution is heavily asymmetric with the probability of the first mode being 0.9, which entails that the majority of consumers are highly sensitive to price. This last feature is completely absent in the logit and Gaussian specifications, which seem to capture only the first mode of the distribution as we can see in Figure 6. Once again the homogeneous parameters are relatively unchanged with respect to the previous specifications. The Gaussian RC on the gasoline still displays a high variance (the standard deviation of the RC equals 2.8).

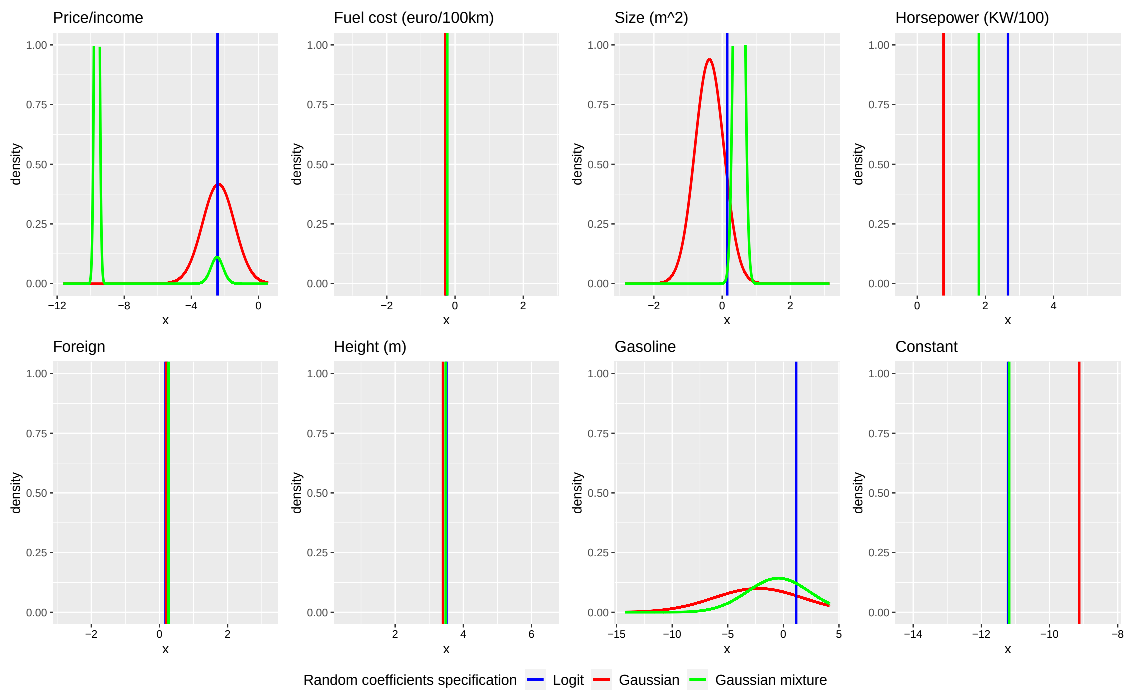
Table 10: Estimation Gaussian mixture on Price

	Baseline		**× Income (/1000)**		**× Pop. density(/100)**		**× Time trend**	
Homogeneous Preferences	$\hat{\beta}$	S.E	$\hat{\beta}$	S.E	$\hat{\beta}$	S.E	$\hat{\beta}$	S.E
Price/income	-	-	-	-	-	-	-	-
Fuel Cost	-0.23	5.8e-03	-	-	-	-	0.026	1e-03
Size(m^2)	-	-	-0.0055	3.7e-04	-	-	-	-
Horsepower(KW/100)	1.8	3.6e-02	-0.0016	1.1e-03	-	-	-0.1	7e-03
Foreign	0.26	6.1e-02	-0.021	1.2e-03	-	-	-	-
Height(m)	3.5	1.1e-02	-0.012	1.2e-03	-0.032	3.7e-04	-	-
Gasoline	-	-	-0.026	1.3e-03	-	-	-	-
Gaussian RC	$\hat{\beta}$	S.E	$\hat{\sigma}$	S.E				
Size(m^2)	0.5	1.9e-02	0.1	6.7e-02	-	-	-	-
Gasoline	-0.45	3.8e-03	2.8	9.1e-03	-	-	-	-
Gaussian Mixture	$\hat{\beta}_1$	S.E	$\hat{\sigma}_1$	S.E	$\hat{\beta}_2$	S.E	$\hat{\sigma}_2$	S.E
Price/income	-9.6	1.8e-02	0.1	1.8e-03	-2.5	1.8e-02	0.35	5.2e-04
Probability	0.9	6.8e-05						

Note: Brand, Year and State FE's are included.

In Figure 6, we plot the estimated distribution of random coefficients under the three specifications we consider. We observe little to no variation in the homogeneous parameters from one specification to the other. The main difference comes from the introduction of the Gaussian mixture on price, which reveals the presence of a large group of highly price sensitive consumers.

Figure 6: Estimated distributions of RCs in the three specifications



Specification test By increasing the flexibility on the distribution of RCs, we recover less precise estimates and the model becomes more difficult to estimate. Thus, it is important to show that the additional flexibility substantially reduces the misspecification of the model. To quantify the degree of misspecification accross the different models, we keep the same set of estimation instruments across the different specifications of RCs and we report the value of the associated Sargan-Hansen J statistics in each case. Moreover, for every model, we follow the procedure developed in section 5 to test if the distribution of RCs on price is well specified. We use the global interval instruments and we denote this test “Interval test”. We report the values of the test statistics and the degrees of freedom of the chi-square under the null in Table 11. We observe an important decrease in the Sargan-Hansen J statistic when we transition from the logit to the Gaussian RC. However, the decrease in the Sargan-Hansen J statistic is much larger when we transition from the Gaussian RC on price to the Gaussian mixture, which indicates that the Gaussian mixture performs much better than the simple Gaussian at capturing the underlying heterogeneity in price sensitivity. The interval test displays a similar behavior, with the largest decrease in the test statistic stemming from the transition from the Gaussian RC to the Gaussian mixture.

Table 11: Evolution of misspecification with flexibility

Instruments	Logit			Gaussian RC			Gaussian mixture		
Test	Stat.	Critical val.	DF	Stat.	Critical val.	DF	Stat.	Critical val.	DF
J test	2755.7	40.1	27	2341.7	36.4	24	950.3	33.9	21
Interval test	1331.9	14.1	7	999.4	14.1	7	244.0	14.1	7

7.4 Counterfactual quantities

The objective of this subsection is to illustrate how changes in the distribution of the RC associated with price affect many counterfactual quantities of interest in IO, such as the price elasticities, the marginal costs faced by car manufacturers, and the pass-through of cost. In order to compare our empirical results with the findings in Miravete et al. (2022), we also derive the demand curvature under the different specifications. They show that a large demand curvature is necessary to recover a pass-through larger than one. We now define these different quantities and derive them under the different specifications considered previously. For exposition purposes, we omit the dependence of the market shares in δ_t, x_{2t} and f, and simply write $s_j(\mathbf{p})$ instead of $\rho_j(\delta_t, x_{2t}; f)$, where $\mathbf{p}$ is the price vector. In Appendix E, we provide analytical formulas for every quantity of interest. The quantities of interest are computed using the year 2018, which is the last year of our sample.

- The price elasticity of demand is the ratio of the percentage change in quantity demanded of a product to the percentage change in price. The price elasticity for product j writes as follows: $\eta_j^1(\mathbf{p}) \equiv \frac{p_j}{s_j}\frac{\partial s_j(\mathbf{p})}{\partial p_j}$.

- The demand curvature of the demand function is given by: $\eta_j^2(\mathbf{p}) \equiv \frac{\partial^2 s_j(\mathbf{p})}{\partial p_j^2}\left(\frac{\partial s_j(\mathbf{p})}{\partial p_j}\right)^{-2}$.

- Marginal costs and mark-ups. To recover the marginal costs and the implied mark-ups, we need to make additional assumptions on the supply side. Following the literature, we consider that each multi-product firm $f \in F$ sets prices for its own products in accordance with a Bertrand-Nash equilibrium. The profit of each firm writes:

  $\Pi_f(\mathbf{p}) = \sum_t \sum_{j \in J_f} (p_j - c_j) M_t s_{jt}(\mathbf{p})$,

  where J_f is the set of goods produced by firm f, c_j is the marginal cost for good j, M_t is the market size and $s_j(p)$ is the market share of product j. The first-order condition with respect to price p_j writes:

  $$\sum_t M_t \; s_{jt}(\mathbf{p}) + \sum_t M_t \sum_{j' \in J_f} (p_{j'} - c_{j'}) \frac{\partial s_{j'}(\mathbf{p})}{\partial p_j} = 0.$$

  We gather all the FOCs and rewrite them in matricial form:

  $$\mathbf{s}(\mathbf{p}) + (\boldsymbol{\Delta}(\mathbf{p}))(\mathbf{p} - \mathbf{c}) = 0.$$

  where $\boldsymbol{\Delta}(\mathbf{p}) = \sum_t M_t \frac{\partial s_{j'}(\mathbf{p})}{\partial p_j}$ if j' and j are produced by the same firm and equals to zero otherwise. $\boldsymbol{\Delta}(\mathbf{p})$ is known as the ownership matrix. Assuming that the prices are in equilibrium, one can recover the marginal costs using the following equation:

  $$\mathbf{c} = \mathbf{p} - (\Delta(\mathbf{p}))^{-1} \mathbf{s}(\mathbf{p}).$$

  The mark-up for product j simply writes: $p_j - c_j$.

- The pass-through of cost is defined as follows. Let us assume that the marginal cost for product j goes from c_j to c'_j (with $c'_j > c_j$), then the cost pass-through equals $\alpha_j = \frac{p'_j - p_j}{c'_j - c_j}$, where p'_j is the new equilibrium price. The pass-through corresponds to the proportion of the cost increase that is transmitted to the price. Following the literature, we derive the pass-through by increasing the marginal costs of each product by 1% and recomputing the marginal cost.

Summary of results We report the median values for the five counterfactual quantities of interest in Table 12. Several remarks are in order. First, the Gaussian mixture yields a much lower price elasticity than the two other specifications. This is related to the emergence of a group of very price sensitive consumers in the mixture specification, which we fail to detect with the logit and Gaussian RC specifications. Moreover, the low price elasticities that we recover in the Gaussian and logit specifications, generate unreasonably low marginal costs (even negative ones as we can see in Figure 7) and excessive mark-ups. In contrast, this problem does not appear with the Gaussian mixture. Finally, to link our results with the findings in Miravete et al. (2022), we now focus on the demand curvature and the pass-through of cost. As expected, the logit displays a curvature and a pass-through equal to 1. In contrast, we can see that the Gaussian mixture displays a larger demand curvature than the other two specifications. This comes from the skewness that the mixture induces in the distribution of price sensitivity. This last feature implies that the Gaussian mixture yields a pass-through much greater than 1 (1.5 on average). Unfortunately, the negative marginal costs we recover with the Gaussian RC prevent us from computing the pass-through in this case.[34]

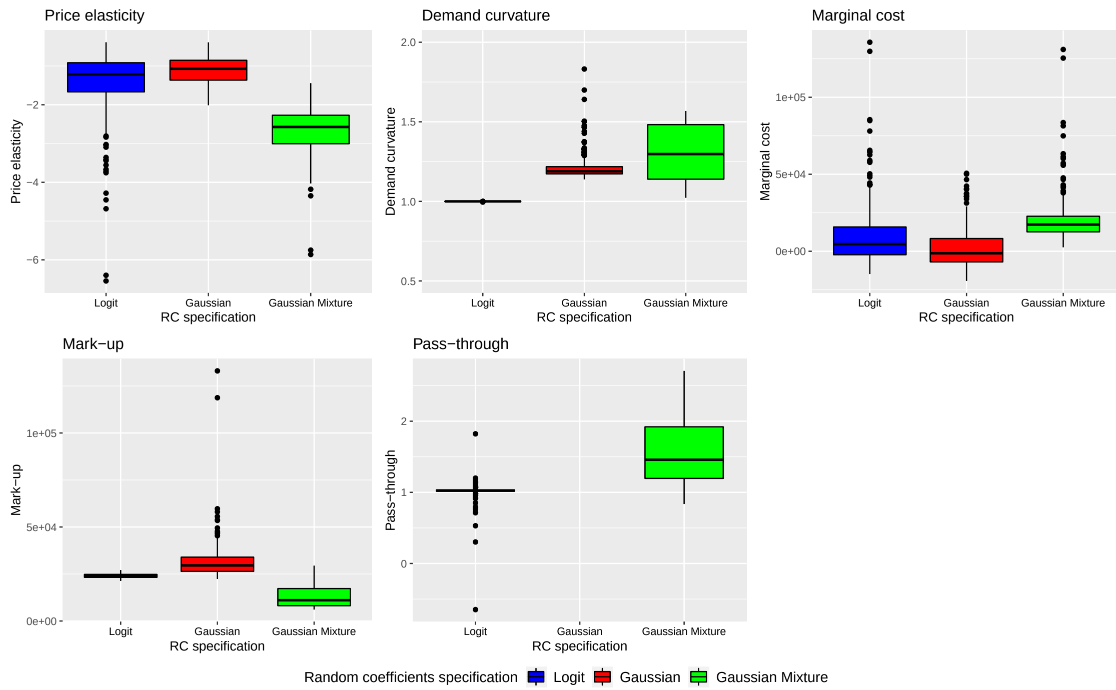
Table 12: Median counterfactual quantities under different specifications on RCs

RC distribution on price	Logit	Gaussian	Gaussian Mixture
Own price-elasticity	-1.2	-1.1	-2.6
Demand curvature	1.0	1.2	1.3
Marginal cost	9,366	1,929	20,105
Mark-up	24,048	29,572	11,066
Pass-through	1.0	-	1.5

In Figure 7, we plot the empirical distributions of the counterfactual quantities. We can see in the plot featuring the distribution of marginal costs that the logit and Gaussian specifications generate negative marginal costs for some of the cars. This is an indication that the price elasticities implied by these specifications are too low in absolute value.

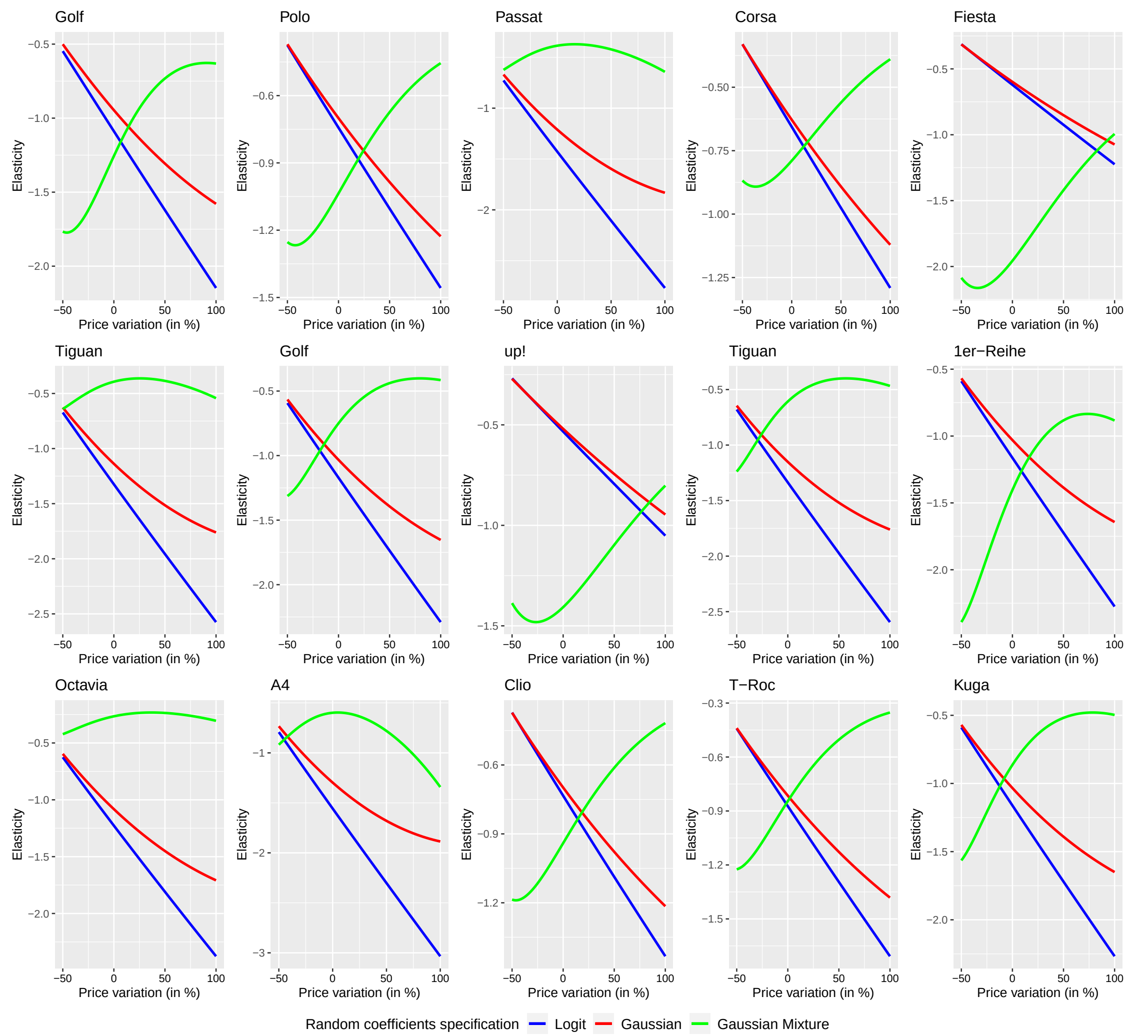
[34]Our algorithm to compute the new equilibrium prices after the change in cost does not converge.

Figure 7: Empirical distribution of counterfactual quantities under different specifications



Finally, in Figure 8, we plot the elasticity functions implied by the different specifications for the 15 most popular cars in our sample. We observe important differences in the elasticities. The Gaussian mixture generates lower price elasticities than the other two specifications. We do the same exercise with the demand curves in Appendix E.

Figure 8: Estimated elasticities under different specifications



8 Conclusion

In this paper, we develop novel econometric tools to parsimoniously increase the flexibility of the distribution of random coefficients in the BLP demand model initiated by Berry et al. (1995). Specifically, we construct novel instruments designed to detect deviations from the true distribution of random coefficients. Building on these instruments, we provide a formal moment-based specification test on the distribution of random coefficients, which allows researchers to test the chosen specification without having to re-estimate the model under a more flexible parametrization. Our instruments are designed to maximize the power of the test when the distribution of RC is misspecified. By exploiting the duality between estimation and testing, we show that these instruments can also improve the estimation of the BLP model under a flexible parametrization. Our Monte Carlo simulations confirm that the interval instruments we develop in this paper outperform the traditional instruments both for testing and estimating purposes. Finally, we apply these new tools to flexibly estimate the demand for cars in Germany. We show that these tools can be applied to the equally popular mixed logit demand model with individual-level data.

In future works, we plan to see if we can generalize these instruments to other non-linear moment-based models, as well as to the general problem of testing distributional assumptions in structural models. From a broader perspective, our paper is part of an existent discussion on the most effective way to model unobserved preference heterogeneity in structural models. Most empirical frameworks feature a clear trade-off between the degree of flexibility one chooses and the precision of the estimates one obtains. It is thus critical to understand how misspecification on the unobserved heterogeneity affects the counterfactual quantities of interest. In the case of the BLP demand model, our paper and others show that misspecification in the distribution of random coefficients substantially distorts the substitution patterns as well as the shape of the demand curve and, thus, is likely to significantly alter the counterfactual quantities.

Bibliography

Alberini, A. and M. Horvath (2021): "All car taxes are not created equal: evidence from Germany," *Energy Economics*, 100.

Allcott, H. and N. Wozny (2014): "Gasoline prices, fuel economy, and the energy paradox," *Review of Economics and Statistics*, 96, 779–795.

Allen, R. and J. Rehbeck (2020): "Identification of random coefficient latent utility models," *arXiv:2003.00276*.

Amemiya, T. (1974): "The nonlinear two-stage least-squares estimator," *Journal of Econometrics*, 2, 105–110.

Andrews, D. W. and P. Guggenberger (2019): "Identification and singularity-robust inference for moment condition models," *Quantitative Economics*, 10, 1703–1746.

Armstrong, T. B. (2016): "Large market asymptotics for differentiated product demand estimators with economic models of supply," *Econometrica*, 84, 1961–1980.

Bahadur, R. R. (1960): "Stochastic comparison of tests," *The Annals of Mathematical Statistics*, 31, 276–295.

Barahona, N., C. Otero, S. Otero, and J. Kim (2020): "Equilibrium effects of food labeling policies," *SSRN working paper*.

Berry, S. T. (1994): "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 25, 242–262.

Berry, S. T. and P. A. Haile (2009): "Nonparametric identification of multinomial choice demand models with heterogeneous consumers," *NBER working paper 15276*.

——— (2014): "Identification in differentiated products markets using market level data," *Econometrica*, 82, 1749–1797.

Berry, S. T., J. Levinsohn, and A. Pakes (1995): "Automobile prices in market equilibrium," *Econometrica*, 63, 841–890.

Berry, S. T., O. B. Linton, and A. Pakes (2004): "Limit theorems for estimating the parameters of differentiated product demand systems," *The Review of Economic Studies*, 71, 613–654.

Chamberlain, G. (1987): "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 34, 305–334.

Chetverikov, D. and D. Wilhelm (2017): "Nonparametric instrumental variable estimation under monotonicity," *Econometrica*, 85, 1303–1320.

Compiani, G. (2018): "Nonparametric demand estimation in differentiated products markets," *SSRN working paper*.

Conlon, C. and J. Gortmaker (2020): "Best practices for differentiated products demand estimation with pyblp," *The RAND Journal of Economics*, 51, 1108–1161.

Crawford, G. S., O. Shcherbakov, and M. Shum (2019): "Quality overprovision in cable television markets," *American Economic Review*, 109, 956–995.

D'Haultfœuille, X., P. Givord, and X. Boutin (2014): "The environmental effect of green taxation: the case of the French *bonus/malus*," *The Economic Journal*, 124, F444–F480.

Dubé, J.-P., J. T. Fox, and C.-L. Su (2012): "Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation," *Econometrica*, 80, 2231–2267.

Dubois, P., R. Griffith, and M. O'Connell (2018): "The effects of banning advertising in junk food markets," *The Review of Economic Studies*, 85, 396–436.

Dufour, J.-M. and M. L. King (1991): "Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or nonstationary AR (1) errors," *Journal of Econometrics*, 47, 115–143.

Dunker, F., S. Hoderlein, and H. Kaido (2022): "Nonparametric identification of random coefficients in endogenous and heterogeneous aggregate demand models," *arXiv preprint: 2201.06140.*

Durrmeyer, I. (2022): "Winners and losers: the distributional effects of the French feebate on the automobile market," *The Economic Journal*, 132, 1414–1448.

Durrmeyer, I. and M. Samano (2018): "To rebate or not to rebate: fuel economy standards vs. feebates," *The Economic Journal*, 128, 3076–3116.

Fox, J. T. and A. Gandhi (2011): "Identifying demand with multidimensional unobservables: a random functions approach," *NBER working paper 17557.*

Fox, J. T., K. il Kim, S. P. Ryan, and P. Bajari (2012): "The random coefficients logit model is identified," *Journal of Econometrics*, 166, 204–212.

Freyberger, J. (2015): "Asymptotic theory for differentiated products demand models with many markets," *Journal of Econometrics*, 185, 162–181.

Gandhi, A. and J.-F. Houde (2019): "Measuring substitution patterns in differentiated-products industries," *NBER working paper 26375.*

Geweke, J. (1981): "The approximate slopes of econometric tests," *Econometrica*, 49, 1427–1442.

Gillingham, K. and S. Houde (2021): "Consumer myopia in vehicle purchases: evidence from a natural experiment," *American Economic Journal: Economic Policy*, 207–238.

Gourieroux, C. and A. Monfort (1995): *Statistics and econometric models*, vol. 2, Cambridge University Press.

Grennan, M. (2013): "Price discrimination and bargaining: empirical evidence from medical devices," *American Economic Review*, 103, 145–177.

Grieco, P. L., C. Murry, and A. Yurukoglu (2022): "The evolution of market power in the US automobile industry," *NBER working paper 29013.*

Grigolon, L., M. Reynaert, and F. Verboven (2018): "Consumer valuation of fuel costs and tax policy: evidence from the European car market," *American Economic Journal: Economic Policy*, 10, 193–225.

Hansen, L. P. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica*, 50, 1029–1054.

Ho, K. and A. Pakes (2014): "Hospital choices, hospital prices, and financial incentives to physicians," *American Economic Review*, 104, 3841–3884.

Huse, C. and N. Koptyug (2022): "Salience and policy instruments: evidence from the auto market," *Journal of the Association of Environmental and Resource Economists*, 9, 345–382.

Ichimura, H. and T. S. Thompson (1998): "Maximum likelihood estimation of a binary choice model with random coefficients of unknown distribution," *Journal of Econometrics*, 86, 269–295.

Ketz, P. (2019): "On asymptotic size distortions in the random coefficients logit model," *Journal of Econometrics*, 212, 413–432.

Kunert, U. (2018): "Diesel fuel and passenger cars receive preferential tax treatment in Europe: reform of taxation needed in Germany," *DIW Weekly Report*, 8, 289–298.

Lee, J. and K. Seo (2015): "A computationally fast estimator for random coefficients logit demand models using aggregate data," *The RAND Journal of Economics*, 46, 86–102.

Lewbel, A. (2000): "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables," *Journal of Econometrics*, 97, 145–177.

Lu, Z., X. Shi, and J. Tao (2021): "Semi-nonparametric estimation of random coefficient logit model for aggregate demand," *SSRN working paper*.

McFadden, D. and K. Train (2000): "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, 15, 447–470.

Miller, N. H., G. Sheu, and M. C. Weinberg (2021): "Oligopolistic price leadership and mergers: the United States beer industry," *American Economic Review*, 111, 3123–3159.

Miller, N. H. and M. C. Weinberg (2017): "Understanding the price effects of the MillerCoors joint venture," *Econometrica*, 85, 1763–1791.

Miravete, E. J., M. J. Moral, and J. Thurk (2018): "Fuel taxation, emissions policy, and competitive advantage in the diffusion of European diesel automobiles," *The RAND Journal of Economics*, 49, 504–540.

Miravete, E. J., K. Seim, and J. Thurk (2022): "Robust pass-through estimation in discrete choice models," *Working paper*.

Nevo, A. (2000): "Mergers with differentiated products: the case of the ready-to-eat cereal industry," *The RAND Journal of Economics*, 31, 395–421.

Newey, W. K. (1990): "Efficient instrumental variables estimation of nonlinear models," *Econometrica*, 58, 809–837.

——— (2004): "Efficient semiparametric estimation via moment restrictions," *Econometrica*, 72, 1877–1897.

Petrin, A. (2002): "Quantifying the benefits of new products: the case of the minivan," *Journal of Political Economy*, 110, 705–729.

Reynaert, M. (2021): "Abatement strategies and the cost of environmental regulation: emission standards on the European car market," *The Review of Economic Studies*, 88, 454–488.

Reynaert, M. and F. Verboven (2014): "Improving the performance of random coefficients demand models: the role of optimal instruments," *Journal of Econometrics*, 179, 83–98.

Roodman, D. (2009): "A Note on the theme of too many instruments," *Oxford Bulletin of Economics and Statistics*, 71, 135–158.

Salanié, B. and F. A. Wolak (2019): "Fast, "robust", and approximately correct: estimating mixed demand systems," *NBER working paper 25726*.

Silvey, S. D. (1959): "The Lagrangian multiplier test," *The Annals of Mathematical Statistics*, 30, 389–407.

Tebaldi, P., A. Torgovitsky, and H. Yang (2019): "Nonparametric estimates of demand in the California health insurance exchange," *NBER working paper 25827*.

Train, K. E. (2008): "EM algorithms for nonparametric estimation of mixing distributions," *Journal of Choice Modelling*, 1, 40–69.

Wang, A. (2022): "Sieve BLP: A semi-nonparametric model of demand for differentiated products," *Journal of Econometrics*.

White, H. (1982): "Maximum likelihood estimation of misspecified models," *Econometrica*, 50, 1–25.

A Extension to the mixed logit demand model

The main difference between the BLP demand model and the mixed logit model is that the latter one assumes that the econometrician observes individual data. Let us consider the baseline mixed logit model with no endogeneity and consumer level data.[35] Indirect utility function of consumer i making choice $j \in \{0, 1, \dots, J\}$ is given by:

[35] In the mixed logit case, the absence of endogenous variables here is not an unrealistic assumption as the econometrician can always model unobserved product quality by incorporating product fixed effects into the utility function

$$u_{ij} = x'_{1ij}\beta_0 + x'_{2ij}v_i + \varepsilon_{ij}, \tag{A.12}$$

where

- ε_{ij} is a preference shock that follows a type I extreme value distribution independent of all other variables and across i, j;
- x_{1ij} is a vector of product characteristics interacted with consumer characteristics of size K_1 which display no preference heterogeneity;
- x_{2ij} is a vector of product characteristics interacted with consumer characteristics of size K_2 which display preference heterogeneity;
- v_i is a vector of random coefficients of size K_2 which jointly follows a joint distribution characterized by a density f;

Each consumer chooses the product that maximizes his or her utility in each market. For any couple $(\tilde{f}, \tilde{\beta})$, demand for product j from consumer i writes:

$$\forall j \neq 0, \quad \rho_j(x_i, \tilde{\beta}, \tilde{f}) = \int_{\mathbb{R}^{K_2}} \frac{\exp(x'_{1ij}\tilde{\beta} + x'_{2ij}v)}{1 + \sum_{k=1}^{J} \exp\left\{x'_{1ik}\tilde{\beta} + x'_{2ik}v\right\}} \tilde{f}(v)dv.$$

For the outside option, we have:

$$\text{for } j = 0, \quad \rho_j(x_i, \tilde{\beta}, \tilde{f}) = \int_{\mathbb{R}^{K_2}} \frac{1}{1 + \sum_{k=1}^{J} \exp\left\{x'_{1ik}\tilde{\beta} + x'_{2ik}v\right\}} \tilde{f}(v)dv.$$

Structural error As we did in the case of the BLP demand model, we can define the structural error generated by $(\tilde{\beta}, \tilde{f})$ as follows. Let y_{ij} equal to 1 if individual i chooses good $j = 0, 1, \dots, J$.

$$\xi_{ij}(\tilde{\beta}, \tilde{f}) = y_{ij} - \rho_j(x_i, \tilde{\beta}, \tilde{f})$$

By construction, at the true (f, β), we have $\mathbb{E}[\xi_{ij}(\beta, f)|x_i] = \mathbb{E}[y_{ij}|x_i] - \rho_j(x_i, \beta, f) = 0$a.s..

Most powerful instrument and approximations As in the aggregate demand model, let us see how we can construct instruments to detect misspecification in the distribution of RC. Given that the model displays no endogeneity, the set of exogenous variables is simply x_i. We now want to find the transformation of x_i which provides the most detection power against a wrong distribution. With this objective in mind we consider a situation where the econometrician has a candidate (f_0, β_0) and wants to test that the model is well specified, namely: $H_0 : (f, \beta) = (f_0, \beta_0)$. Under an alternative $H_1 : (f, \beta) = (f_a, \beta_a)$, the expression for the Most Powerful Instrument (i.e the instrument which maximizes the correlation between the Structural Error and any instrument in the class of measurable functions of x_i) is the same as previously:

$$\begin{aligned}\mathbb{E}[\Delta_{0,a}^{\xi_j}|x_{ij}] &= \Delta_j(x_i, f_0, \beta_0, f_a, \beta_a) \\ &= \rho_j(x_i, \beta_0, f_0) - \rho_j(x_i, \beta_a, f_a) \\ &= \int_{\mathbb{R}} \rho_j(x_i, \beta_0, f_0) - \frac{\exp(x'_{1ij}\beta_a + x'_{2ij}v)}{1 + \sum_{k=1}^{J} \exp\{x'_{1ik}\beta_a + x'_{2ik}v\}} f_a(v)\end{aligned}$$

Several remarks are in order. First, contrary to the BLP case, the correction term $\Delta_{0,a}^{\xi_j}$ is a function of the exogenous variables x_i and thus we don't need to estimate its conditional expectation.Second, β_a and f_a are usually unknown to the econometrician and thus we cannot exploit directly this expression. As did for the BLP case, we propose 2 feasible approximations of the MPI.

- **Global approximation:** we replace the unknown β_a by a known substitute β_0[36]. As for the unknown distribution of RC f_a, we proceed as in the BLP case and we replace the integral with a finite sum. Namely, we have:

$$\mathbb{E}[\Delta_j(x_i, f_0, f_a)|x_i] \approx \sum_{l=1}^{L} \omega_l \underbrace{\left[\rho_j(x_i, \beta_0, f_0) - \frac{\exp(x'_{1ij}\beta_0 + x'_{2ij}v_l)}{1 + \sum_{k=1}^{J} \exp\{x'_{1ik}\beta_0 + x'_{2ik}v_l\}}\right]}_{\pi_{j,l}(x_i)}$$

  with $\{v_l\}_{l=1,\ldots,L}$ L points chosen in the support of f_a, and ω_l the unknown weights associated with each point

- **Local approximation:** we provide a local approximation which is accurate when f_0 is close to the true density f_a. To derive this local approximation, we need to impose additional restrictions on β_0 and β_a so that $\|\beta_a - \beta_0\| = O\left(\int_{\mathbb{R}^{K_2}} |f_0(v) - f_a(v)|dv\right)$

  Assumption 1 *We assume that $\beta_0 = \beta_0^*$ and $\beta_a = \beta_a^*$ where (β_0^*, β_a^*) are both pseudo true values which maximize the conditional expectation of their respective*

[36]in simulations, we find that the homogeneous parameters are usually close to each other even when the distributions are somewhat remote from each other

population log-likelihoods. Namely,

$$\beta_0^* = \underset{\beta \in \mathbb{R}^{K_1}}{argmax}\, \mathbb{E}\big[L(x_i, y_i, \beta, f_0)\big|x_i\big] \text{ with } L(x_i, y_i, \beta, f_0) = \sum_{j=0}^{J} \mathbf{1}\{y_{ij} = 1\} \log(\rho_j(x_i, \beta, f_0))$$

$$\beta_a^* = \underset{\beta \in \mathbb{R}^{K_1}}{argmax}\, \mathbb{E}\big[L(x_i, y_i, \beta, f_a)\big|x_i\big] \text{ with } L(x_i, y_i, \beta, f_a) = \sum_{j=0}^{J} \mathbf{1}\{y_{ij} = 1\} \log(\rho_j(x_i, \beta, f_a))$$

Now we can derive the following first order approximation of the $\Delta_j(x_i, f_0, \beta_0, f_a, \beta_a)$

Proposition 1.1
Under Assumption1, a first order expansion of $\Delta_j(x_i, f_0, \beta_0, f_a, \beta_a)$ around f_0 writes:

$$\begin{aligned}
\Delta_j(x_i, f_0, \beta_0, f_a, \beta_a) &= g_j(x_i, \beta_0, f_0) - g_j(x_i, \beta_a, f_a) \\
&= \int_{\mathbb{R}^{K_2}} \frac{exp(x'_{1ij}\beta_0 + x'_{2ij}v)}{1 + \sum_{k=1}^{J} \exp\{x'_{1ik}\beta_0 + x'_{2ik}v\}} (f_0(v) - f_a(v))dv + \frac{\partial \rho_j(x_i, \beta, f_a)}{\partial \beta}\bigg|_{\beta=\beta_0} (\beta_1 - \beta_0) + \mathcal{R}_0
\end{aligned}$$

with $\mathcal{R}_0 = \int_{\mathbb{R}^{K_2}} |f_0(v) - f_a(v)|dv$

The proof is in section B. Building on this approximation, we can construct the following local feasible approximation of the MPI:

$$\mathbb{E}[\Delta_j(x_i, f_0, f_a)|x_i] \approx \sum_{l=1}^{L} \bar{\omega}_{1l} \underbrace{\left[\rho_j(x_i, \beta_0, f_0) - \frac{\exp(x'_{1ij}\beta_0 + x'_{2ij}v_l)}{1 + \sum_{k=1}^{J} \exp\{x'_{1ik}\beta_0 + x'_{2ik}v_l\}}\right]}_{\bar{\pi}_{1,j,l}(x_i)}$$

$$+ \sum_{l=1}^{L} \bar{\omega}_{2l} \underbrace{\frac{\partial}{\partial \beta}\left\{\frac{exp(x'_{1ij}\beta_0 + x'_{2ij}v_l)}{1 + \sum_{k=1}^{J} \exp\{x'_{1ik}\beta_0 + x'_{2ik}v_l\}}\right\}}_{\bar{\pi}_{2,j,l}(x_i)}$$

with $\{v_l\}_{l=1,\dots,L}$ L points chosen in the support of f_a, and $\bar{\omega}_l$ the unknown weights associated with each point. The interval instruments are simply the set $(\bar{\pi}_{1,j,l}(x_i), \bar{\pi}_{2,j,l}(x_i))$.

Specification test

B Proofs

B.1 Identification

In this subsection, we prove that under Assumption A, the distribution of random coefficients f is non-parametrically point identified.

B.1.1 Proof of Proposition 2.1

We want to show that under Assumptions A, the following implication holds:

$$\begin{aligned}(\tilde{f},\tilde{\beta})=(f,\beta) &\iff \mathbb{E}[\xi_{jt}(\tilde{f},\tilde{\beta})|z_{jt}]=0 \ a.s.\\ &\iff \mathbb{E}\left[\rho_j^{-1}(s_t,x_{2t},\tilde{f})-x_{1jt}'\tilde{\beta}\middle|z_{jt}\right]=0 \ a.s..\end{aligned}$$

Step 1 First, we show that for any random permutation of indexes $j \to j'$, the following equivalence holds:

$$\mathbb{E}[\xi_{jt}|z_{jt}]=0 \ a.s. \iff \mathbb{E}[\xi_{jt}|z_{j't}]=0 \ a.s. \ \forall \ j'.$$

First, let us show that the standard exogeneity conditions assumed in Berry and Haile (2014) and in Wang (2022) implies the moment condition we utilize in this paper:

By construction, we can rewrite the exogeneity condition A (i) as follows:

$$\mathbb{E}[\xi_{jt}|z_{jt}]=\sum_{k=1}^{J}\Pr(j=k)\mathbb{E}\left[\xi_{jt}\middle|z_{jt},j=k\right]=\frac{1}{J}\sum_{k=1}^{J}\mathbb{E}\left[\xi_{jt}\middle|z_{jt},j=k\right]$$

The exogeneity condition in Wang (2022) assumes: $\forall k, \quad \mathbb{E}\left[\xi_{jt}\middle|z_{jt},j=k\right]=0$. From what precedes, this condition implies the exogeneity condition $\mathbb{E}[\xi_{jt}|z_{jt}]=0 \ a.s.$ in A (i). This assumption is required for non-parametric identification of the demand functions but not for the non-parametric identification of the distribution of RC.

Now let us prove the identification result. As an artifact for our proof, let us consider a new indexation, which is done exogenously across markets. We denote j' the exogenous indices. Consequently, a same product j doesn't necessarily have the same indices across markets. As the new indexation is done exogenously, we have for any j':

$$\mathbb{E}[\xi_{jt}(\tilde{f},\tilde{\beta})|z_{jt}]=\mathbb{E}[\xi_{jt}(\tilde{f},\tilde{\beta})|z_{jt},j\to j']\equiv\mathbb{E}_{j'}[\xi_{j't}(\tilde{f},\tilde{\beta})|z_{j't}]a.s.$$

$j \to j'$ indicates index j has been changed into j'. Consequently, we have:

$$\mathbb{E}[\xi_{jt}(\tilde{f},\tilde{\beta})|z_{jt}]=0 \ a.s. \iff \forall j' \ \mathbb{E}_{j'}[\xi_{j't}(\tilde{f},\tilde{\beta})|z_{j't}]=0 \ a.s.$$

As a consequence, we can rewrite the initial equivalence as follows:

$$(\tilde{f},\tilde{\beta})=(f,\beta) \iff \forall j', \ \mathbb{E}_{j'}[\xi_{j't}(\tilde{f},\tilde{\beta})|z_{j't}]=0 \ a.s.$$

Given the random permutation $j \to j'$, which is market dependent, we must redefine our matrices and vectors as follows: $\hat{x}_t = M_t x_t$ with $(M_t)_{i,k} = \mathbf{1}\{i = j_t, k = j'_t\}$. Likewise $\hat{s}_t = M_t s_t$. M_t is a random matrix. It is straight forward to show the direct implication.

$$(\tilde{f}, \tilde{\beta}) = (f, \beta) \implies \quad \forall j', \qquad \mathop{\mathbb{E}}_{j'}\left[\rho_{j'}^{-1}(\hat{s}_t, \hat{x}_{2t}, \tilde{f}) - x'_{1j't}\tilde{\beta}\middle| z_{j't}\right] = \mathop{\mathbb{E}}_{j'}[\xi_{j't}(f, \beta)|z_{j't}] = 0 \ a.s.$$

The reverse implication is much more intricate to prove and we will exploit other results in the literature. We want to show:

$$(\tilde{f}, \tilde{\beta}) \neq (f, \beta) \implies \quad \exists j' \ \Bigg| \ \mathop{\mathbb{E}}_{j'}\left[\rho_{j'}^{-1}(\hat{s}_t, \hat{x}_{2t}, \tilde{f}) - \hat{x}'_{1jt}\tilde{\beta}\middle| z_{j't}\right] = 0 \text{ a.s. does not hold}$$

First, let us assume that $\tilde{f} = f$ and $\tilde{\beta} \neq \beta$, then we have:

$$\rho^{-1}(\hat{s}_t, \hat{x}_{2t}, \tilde{f}) - \hat{x}_{1t}\tilde{\beta} = \underbrace{\rho^{-1}(\hat{s}_t, \hat{x}_{2t}, f) - x_{1t}\beta}_{\hat{\xi}_t(f,\beta)} + \hat{x}_{1t}(\beta - \tilde{\beta})$$

By assumption, we have: $P(x'_{1t}x_{1t} \ \ dp) > 0$. M_t is symmetric, idempotent and full rank. As a consequence,

$$P(\hat{x}'_{1t}\hat{x}_{1t} \ \ dp) = P(x'_{1t}M_t x_{1t} \ \ dp) = P(x'_{1t}x_{1t} \ \ dp) > 0$$

Therefore, we have $\forall \ \gamma \neq 0 \in \mathbb{R}^K$,

$$\begin{aligned} P(\gamma'\hat{x}'_{1t}\hat{x}_{1t}\gamma > 0) > P(\hat{x}'_{1t}\hat{x}_{1t} \ \ dp) > 0 &\iff P(\|\hat{x}_{1t}\gamma\|^2 > 0) > 0 \\ &\iff P(\hat{x}_{1t}\gamma \neq 0) > 0 \end{aligned}$$

Thus, $\exists j' \mid x'_{1j't}(\beta - \tilde{\beta}) = 0 \ a.s$ does not hold. To conclude, there exists j' such that:

$$\mathop{\mathbb{E}}_{j'}[\rho_{j'}^{-1}(\hat{s}_t, \hat{x}_{2t}, f) - x'_{1j't}\tilde{\beta}|z_{j't}] = \underbrace{\mathop{\mathbb{E}}_{j'}[\xi_{j't}(f, \beta)|z_{j't}]}_{=0} + \underbrace{\mathop{\mathbb{E}}_{j'}[x'_{1j't}(\beta - \tilde{\beta})|z_{j't}]}_{= \ 0 \ a.s \text{ does not hold from the completeness}}$$

Now let us assume that $\tilde{f} \neq f$ and we want to show that $\forall \tilde{\beta} \in \mathbb{R}^k$, $\exists j'$ such that:

$$\mathop{\mathbb{E}}_{j'}\left[\rho_{j'}^{-1}(\hat{s}_t, \hat{x}_{2t}, \tilde{f}) - x'_{1j't}\tilde{\beta}\middle| z_{jt}\right] = 0 \ a.s \ \text{ does not hold}$$

First, note that $\forall j'$,

$$\underset{j'}{\mathbb{E}}\left[\rho_{j'}^{-1}(\hat{s}_t,\hat{x}_{2t},\tilde{f})-x'_{1j't}\tilde{\beta}\middle|z_{j't}\right]=\underbrace{\underset{j'}{\mathbb{E}}\left[\xi_{j't}(f,\beta)\middle|z_{j't}\right]}_{=0}+\underset{j'}{\mathbb{E}}\left[\rho_{j'}^{-1}(\hat{s}_t,\hat{x}_{2t},\tilde{f})-\rho_{j'}^{-1}(\hat{s}_t,\hat{x}_{2t},f)-x'_{1j't}(\tilde{\beta}-\beta)\middle|z_{j't}\right]$$

Thus, we need to show that $\exists j' \;\Big|\; \underset{j'}{\mathbb{E}}\Big[\rho_{j'}^{-1}(\hat{s}_t,\hat{x}_{2t},\tilde{f}) - \rho_{j'}^{-1}(\hat{s}_t,\hat{x}_{2t},f) - x'_{1j't}(\tilde{\beta} - \beta)\Big] = 0$ a.s doesn't hold. From the completeness condition, a sufficient condition is: $\exists j' \;\Big|\; \rho_{j'}^{-1}(\hat{s}_t,\hat{x}_{2t},\tilde{f})-\rho_{j'}^{-1}(\hat{s}_t,\hat{x}_{2t},f)-x'_{1j't}(\tilde{\beta}-\beta)=0\ a.s$ does not hold. Let $\gamma=(\tilde{\beta}-\beta)$.

By contradiction, it can be easily be shown that $\rho(\hat{\delta}_t,\hat{x}_{2t},f)-\rho(\hat{\delta}_t+\hat{x}_{1t}\gamma,\hat{x}_{2t},\tilde{f})\neq 0 \implies \exists j'\ \rho_{j'}^{-1}(\hat{s}_t,\hat{x}_{2t},\tilde{f})\neq\rho_j^{-1}(\hat{s}_t,\hat{x}_{2t},f)+\gamma' x_{1j't}$. Indeed, assume that $\rho(\hat{\delta}_t,\hat{x}_{2t},f)-\rho(\hat{\delta}_t+\hat{x}_{1t}\gamma,\hat{x}_{2t},\tilde{f})\neq 0$ and $\forall j'\ \rho_{j'}^{-1}(\hat{s}_t,\hat{x}_{2t},\tilde{f})=\rho_{j'}^{-1}(\hat{s}_t,\hat{x}_{2t},f)+\gamma' x_{1j't}$. Then, we have: $\rho(\rho^{-1}(\hat{s}_t,\hat{x}_{2t},\tilde{f}),\hat{x}_{2t},\tilde{f})=\rho(\rho^{-1}(\hat{s}_t,\hat{x}_{2t},f)+\hat{x}_{1t}\gamma,\hat{x}_{2t},\tilde{f})=\rho(\hat{\delta}_t+\hat{x}_{1t}\gamma,\hat{x}_{2t},\tilde{f})\neq\rho(\hat{\delta}_t,\hat{x}_{2t},f)=\hat{s}_t$. Therefore, we have a contradiction.

Thus, the next step is to show that $\forall\gamma,\ \tilde{f}\neq f \implies \rho(\hat{\delta}_t,\hat{x}_{2t},f_0)-\rho(\hat{\delta}_t+\hat{x}_{1t}\gamma,\hat{x}_{2t},f)=0$ a.s does not hold.

To this end, we are going to exploit the identification result shown by Wang (2022). Following the notations in this paper, we define $\mu_i=\hat{x}_{1t}\Gamma+\hat{x}_{2t}v_i=\hat{x}_t\mathbf{v}$ with $\mathbf{v}_i=(\Gamma,v_i)$. Here Γ is a degenerate random variable characterized by constant c such that $P(\Gamma=c)=1$. Let $G_{\mu|\hat{x}_t}$ the distribution of $\mu_i|\hat{x}_t$ under $f^{\dagger}=(c=0,f)$ and $G_{\tilde{\mu}|\hat{x}_t}$ the distribution of $\mu_i|\hat{x}_t$ under $\tilde{f}^{\dagger}=(c=\gamma,\tilde{f})$. The following result is shown in Wang (2022): for any $\hat{\bar{x}}_t\in Supp(\hat{x}_t)$

$$\exists j' \mid \rho_{j'}(\hat{\delta}_t,G_{\mu|\hat{\bar{x}}_t})-\rho_{j'}(\hat{\delta}_t,G_{\tilde{\mu}|\hat{\bar{x}}_t})=0 \text{ on open set } \mathcal{D}\subset\mathbb{R}^J \implies G_{\mu|\hat{\bar{x}}_t}=G_{\tilde{\mu}|\hat{\bar{x}}_t}$$

Note that thanks to the real analytic property of the demand functions ρ, Wang (2022) does not require a full support assumption on $\hat{\delta}_t$

Fix the value of $\hat{x}_t$ as follows: $\hat{x}_t=\bar{M}_t\bar{x}_t=\hat{\bar{x}}_t$.By assumption, there exists $\bar{x}_t\in Supp(x_t)$ such that $\bar{x}'_t\bar{x}_t$ is dp and $\delta_t=\bar{x}_{1t}\beta+\xi_t$ varies on an open set $\bar{\mathcal{D}}$ almost surely.These properties naturally transmit to $\hat{\bar{x}}_t$. The chosen permutation $\bar{M}_t$ doesn't matter. Given the result in Wang (2022), in order to prove that $\rho(\hat{\delta}_t,\hat{x}_{2t},f_0)-\rho(\hat{\delta}_t+\hat{x}_{1t}\gamma,\hat{x}_{2t},f)=0\ a.s$ does not hold, we just need to prove that $\forall\gamma,\ \tilde{f}\neq f\implies G_{\tilde{\mu}|\hat{\bar{x}}_t}\neq G_{\mu|\hat{\bar{x}}_t}$. As the density functions are assumed to be continuous, $\tilde{f}\neq f \implies \exists v^*\in\mathbb{R}^{K_2}\ \ \tilde{F}(v^*)\neq F(v^*)$. Take $x^*=(0_{K_1},\hat{\bar{x}}_{2t}v^*)'=\hat{\bar{x}}_t(0_{K_1},v^*)'$:

$$G_{\mu|\hat{\bar{x}}_t}(x^*) = P(x_t\mathbf{v}_i \le x^*|x_t = \hat{\bar{x}}_t) = P((x_t'x_t)^{-1}x_t'x_t\mathbf{v}_i \le (x_t'x_t)^{-1}x_t'\bar{x}_t(0_{K_1}, v^*)'|x_t = \hat{\bar{x}}_t)$$
$$= (1_{K_1}, P(v_i \le v^*|x_t = \hat{\bar{x}}_t))' = (1_{K_1}, F(v^*))'$$

The last equality comes from independence of v_i and x_t. Likewise, $G_{\tilde{\mu}|\hat{\bar{x}}_t}(x^*) = (1\{\gamma > 0\}, \tilde{F}(v^*))'$

Therefore, $\exists x^*$, $\forall\gamma\ \ G_{\tilde{\mu}|\hat{\bar{x}}_t}(x^*) \neq G_{\mu|\hat{\bar{x}}_t}(x^*)$. Following the result in Wang (2022), we have that for all $\gamma \in \mathbb{R}^{K_1}$, $\rho(\hat{\delta}_t, \hat{x}_{2t}, f) - \rho(\hat{\delta}_t + \hat{x}_{1t}\gamma, \hat{x}_{2t}, \tilde{f}) = 0\ a.s$ does not hold which in turn implies that for all $\gamma \in \mathbb{R}^{K_1}$, $\exists j'\ \ \rho_j^{-1}(\hat{s}_t, \hat{x}_{2t}, \tilde{f}) - \rho_j^{-1}(\hat{s}_t, \hat{x}_{2t}, f) + \hat{x}_{1jt}'\gamma = 0\ a.s$ does not hold.

To conclude: $\forall\beta \in \mathbb{R}^k$, there exists j' such that:

$$\rho_{j'}^{-1}(\hat{s}_t, \hat{x}_{2t}, \tilde{f}) - \rho_{j'}^{-1}(\hat{s}_t, \hat{x}_{2t}, f) - x_{1j't}'(\tilde{\beta} - \beta) = 0\ a.s\ \text{ does not hold}$$

which is what we wanted to show.

B.1.2 Proof of Corollary ??

Let us assume that specification $\mathcal{F}_0$, instruments $h_E(z_{jt})$ and weighting matrix yields a unique pseudo true value θ_0.

$$\theta_0 = \underset{\tilde{\theta}}{Argmin}\ \mathbb{E}[\xi_{jt}(f_0(.|\tilde{\lambda}, \tilde{\theta})h_E(z_{jt})']W\mathbb{E}[h_E(z_{jt})\xi_{jt}(f_0(.|\tilde{\lambda}, \tilde{\theta})]$$

Under $H_0 : f \in \mathcal{F}_0$ and $f = f_0(\cdot|\lambda)$. By the mean independence assumption on the unobserved quality ξ_{jt}, we have at the true $\theta = (\beta, \lambda)$:

$$\xi_{jt}(f_0(.|\lambda), \beta) = \rho_j^{-1}(s_t, x_{2t}, f_0(.|\lambda)) - x_{1jt}'\beta = \xi_{jt} \implies \mathbb{E}[\big(\xi_{jt}(f_0(.|\lambda), \beta)h_E(z_{jt})] = 0$$

Thus, θ is solution to the previous minimization problem and as the solution is unique: $\theta_0 = \theta$. As a consequence, $\xi_{jt}(f_0(.|\lambda_0), \beta_0)) = \xi_{jt}$ and $\mathbb{E}[\xi_{jt}(f_0(.|\lambda_0), \beta_0)|z_{jt}] = 0\ as$

Under an alternative specification: $f \notin \mathcal{F}_0$, we know from the identification proof that $\forall\tilde{\theta} = (\tilde{\beta}, \tilde{\lambda})$,

$$\mathbb{E}\left[\rho_j^{-1}(s_t, x_{2t}, f_0(.|\tilde{\lambda})) - x_{1jt}'\tilde{\beta}\middle| z_{jt}\right] = 0\ a.s\ \text{ does not hold}$$

In particular, the last equation holds for $\tilde{\theta} = \theta_0$

B.2 Detecting misspecification: the most powerful instrument

Proof of Proposition 3.1.

- Under $\overline{H}_0 : (f, \beta) = (f_0, \beta_0)$. By assumption, the data are i.i.d. across markets, $\mathbb{E}[\|\xi_{jt}(f_0, \beta_0)h_D(z_{jt})\|^2] = \frac{1}{J}\mathbb{E}[\sum_j \|\xi_{jt}(f_0, \beta_0)h_D(z_{jt})\|^2] < +\infty$, the CLT applies:

$$\frac{1}{\sqrt{TJ}} \sum_{j,t} h_D(z_{jt})\xi_{jt}(f_0, \beta_0) = \frac{1}{\sqrt{TJ}} \sum_{j,t} h_D(z_{jt})\xi_{jt} \underset{T\to+\infty}{\longrightarrow} \mathcal{N}(0, \tilde{\Omega}_0),$$

with:

$$\begin{aligned}
\tilde{\Omega}_0 &= \mathbb{E}\left[\left(\frac{1}{\sqrt{J}}\sum_{j=1}^{J} h_D(z_{jt})\xi_{jt}\right)\left(\frac{1}{\sqrt{J}}\sum_{j=1}^{J} h_D(z_{jt})\xi_{jt}\right)'\right] \\
&= \frac{1}{J}\mathbb{E}\left[\sum_{j=1}^{J} h_D(z_{jt})h_D(z_{jt})'\xi_{jt}^2 + \sum_{j=1}^{J}\sum_{k\neq j} h_D(z_{jt})h_D(z_{kt})'\xi_{jt}\xi_{kt}\right] \\
&= \frac{1}{J}\mathbb{E}\left[\sum_{j=1}^{J} h_D(z_{jt})h_D(z_{jt})'\xi_{jt}^2\right] + \frac{1}{J}\sum_{j=1}^{J}\sum_{k\neq j}\mathbb{E}\left[h_D(z_{jt})h_D(z_{kt})'\underbrace{\mathbb{E}[\xi_{jt}\xi_{kt}|z_{jt}, z_{kt}]}_{=0}\right] \\
&= \mathbb{E}\left[h_D(z_{jt})h_D(z_{jt})'\xi_{jt}^2\right] \\
&= \Omega_0.
\end{aligned}$$

Third line comes from $\xi_{jt} \perp\!\!\!\perp \xi_{kt}|z_t$. By assumption, Ω_0 has a full rank. Thus, we have by the CMT:

$$S_T(h_D, f_0, \beta_0) = TJ\left(\frac{1}{TJ}\sum_{j,t}\xi_{jt}(f_0, \beta_0)h_D(z_{jt})\right)'\hat{\Omega}_0^{-1}\left(\frac{1}{TJ}\sum_{j,t}\xi_{jt}(f_0, \beta_0)h_D(z_{jt})\right) \underset{T\to+\infty}{\xrightarrow{d}} \chi^2_{|h_D|_0}.$$

- Under $H_a' :\ \mathbb{E}[h_D(z_{jt})\xi_{jt}(f_0, \beta_0)] \neq 0$. The data are i.i.d. across markets, by the law of large numbers: $\frac{1}{TJ}\sum_{j,t} h_D(z_{jt})\xi_{jt}(f_0, \beta_0) \xrightarrow{P} \mathbb{E}\left[\frac{1}{J}\sum_j h_D(z_{jt})\xi_{jt}(f_0, \beta_0)\right]$. It follows by the continuous mapping theorem:

$$\begin{aligned}
\frac{S_T(h_D, f_0, \beta_0)}{T} &\xrightarrow{P} J\mathbb{E}\left[\frac{1}{J}\sum_j h_D(z_{jt})\xi_{jt}(f_0, \beta_0)\right]'\Omega_0^{-1}\mathbb{E}\left[\frac{1}{J}\sum_j h_D(z_{jt})\xi_{jt}(f_0, \beta_0)\right] \\
&= J\underbrace{\mathbb{E}[h_D(z_{jt})\xi_{jt}(f_0, \beta_0)]'\Omega_0^{-1}\mathbb{E}[h_D(z_{jt})\xi_{jt}(f_0, \beta_0)]}_{\kappa(h_D, f_0, \beta_0)}
\end{aligned}$$

Under H'_a, $\kappa(h_D, f_0, \beta_0)$ is strictly positive because Ω_0 is positive definite. Thence,

$$\begin{aligned}\forall q \in \mathbb{R}, \quad \underset{T\to\infty}{lim}\ \mathbb{P}(S_T(h_D, f_0, \beta_0) > q) &= \underset{T\to\infty}{lim}\ \mathbb{P}\left(\frac{S(h_D, f_0, \beta_0) - q}{T} > 0\right) \\ &= \mathbb{P}(J\kappa(h_D, f_0, \beta_0) > 0) \\ &= 1,\end{aligned}$$

where the second equality holds because convergence in probability implies convergence in distribution.

□

Proof of Proposition 3.2. To shorten notations, let $\xi_{jt0} \equiv \xi_{jt}(f_0(\cdot|\lambda_0), \beta_0)$, $\xi_{jta} \equiv \xi_{jt}(f_a, \beta_a)$ and ξ_{t0} and ξ_{ta} their stacked versions over j. Likewise, we define $h_D(z_t) = (h_D(z_{1t}), ..., h_D(z_{Jt}))'$ The asymptotic slope of the test writes:

$$\begin{aligned}c_{h_D}(f_a, \beta_a) &= \mathbb{E}\left(\sum_j \xi_{jt0} h_D(z_{jt})\right)' \mathbb{E}\left(\left(\sum_j \xi_{jt0} h_D(z_{jt})\right)\left(\sum_{j'} \xi_{j't0} h_D(z_{j't})\right)'\right)^{-1} \mathbb{E}\left(\sum_j \xi_{jt0} h_D(z_{jt})\right) \\ &= \mathbb{E}(\xi'_{t0} h_D(z_t))\mathbb{E}(h_D(z_t)'\xi_{t0}\xi'_{t0} h_D(z_t))^{-1}\mathbb{E}(h_D(z_t)'\xi_{t0}) \\ &= \mathbb{E}({\Delta^{\xi_t}_{0,a}}' h_D(z_t))\mathbb{E}(h_D(z_t)'\mathbb{E}(\xi_{t0}\xi'_{t0}|z_t) h_D(z_t))^{-1}\mathbb{E}(h_D(z_t)'\Delta^{\xi_t}_{0,a})\end{aligned}$$

Third line comes from $\mathbb{E}({\Delta^{\xi_t}_{0,a}}' h_D(z_t)) = \mathbb{E}((\xi_{t0} - \xi_{ta})' h_D(z_t)) = \mathbb{E}(\xi'_{t0} h_D(z_t))$ because ξ_{ta} is the true structural error. Then the slope of the test taking $h^*_D = \mathbb{E}(\xi_{t0}\xi'_{t0}|z_t)^{-1}\mathbb{E}(\Delta^{\xi_t}_{0,a}|z_t)$ is equal to:

$$c_{h^*_D}(f_a, \beta_a) = \mathbb{E}\left(\mathbb{E}(\Delta^{\xi_t}_{0,a}|z_t)'\mathbb{E}(\xi_{t0}\xi'_{t0}|z_t)^{-1}\mathbb{E}(\Delta^{\xi_t}_{0,a}|z_t)\right)$$

To finish the proof, we must show that for any set of instruments h_D, we have: $c_{h^*_D}(f_a, \beta_a) \geq c_{h_D}(f_a, \beta_a)$.

Denote $\tilde{h}_D(z_t) = \mathbb{E}(\xi_{t0}\xi'_{t0}|z_t)^{1/2} h_D(z_t)$ and $\tilde{h}^*_D(z_t) = \mathbb{E}(\xi_{t0}\xi'_{t0}|z_t)^{1/2} h^*_D(z_t)$.With these new notations, we have:

$$\begin{aligned}c_{h^*_D}(f_a, \beta_a) - c_{h_D}(f_a, \beta_a) &= \mathbb{E}\left(\tilde{h}^*_D(z_t)'\tilde{h}^*_D(z_t)\right) - \mathbb{E}\left(\tilde{h}^*_D(z_t)'\tilde{h}_D(z_t)\right)\mathbb{E}\left(\tilde{h}_D(z_t)'\tilde{h}_D(z_t)\right)^{-1}\mathbb{E}\left(\tilde{h}_D(z_t)'h^*_D(z_t)\right) \\ &= G'\begin{pmatrix}\mathbb{E}\left(\tilde{h}^*_D(z_t)'\tilde{h}^*_D(z_t)\right) & \mathbb{E}\left(\tilde{h}^*_D(z_t)'\tilde{h}_D(z_t)\right) \\ \mathbb{E}\left(\tilde{h}_D(z_t)'\tilde{h}^*_D(z_t)\right) & \mathbb{E}\left(\tilde{h}_D(z_t)'\tilde{h}_D(z_t)\right)\end{pmatrix} G \\ &= G'\mathbb{E}\left(\tilde{H}\tilde{H}'\right) G \geq 0\end{aligned}$$

with $\tilde{H} = (\tilde{h}_D^*(z_t), \tilde{h}_D(z_t))'$ and $G = \left(1, -\mathbb{E}\left(\tilde{h}_D^*(z_t)'\tilde{h}_D(z_t)\right)\mathbb{E}\left(\tilde{h}_D(z_t)'\tilde{h}_D(z_t)\right)^{-1}\right)'$

$\square$

Proof of Proposition 3.3.

Under Assumption A, Proposition 2.1 implies the following:

$$\begin{aligned}
\overline{H}_a : (f, \beta) = (f_a, \beta_a) \neq (f_0, \beta_0) &\implies \quad \mathbb{E}[\xi_{jt}(f_0, \beta_0)|z_{jt}] \neq 0 \quad a.s. \\
&\implies \quad \mathbb{E}[\xi_{jt}(f_0, \beta_0)|z_{jt}]^2 > 0 \ a.s. \\
&\implies \quad \mathbb{E}\big[\mathbb{E}[\xi_{jt}(f_0, \beta_0)|z_{jt}]^2\big] > 0 \\
&\implies \quad \mathbb{E}\big[\mathbb{E}[\xi_{jt}(f_0, \beta_0)\mathbb{E}[\xi_{jt}(f_0|z_{jt}]|z_{jt}]\big] > 0 \\
&\implies \quad \mathbb{E}\big[\xi_{jt}(f_0, \beta_0)\mathbb{E}[\xi_{jt}(f_0|z_{jt}]\big] > 0 \\
&\implies \quad \overline{H}'_a : \ \mathbb{E}\big[\xi_{jt}(f_0, \beta_0) \underbrace{\mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]}_{h_D^*(z_{jt})}\big] \neq 0
\end{aligned}$$

Under the same assumptions as 3.1, we have the following:

$$\overline{H}'_a : \ \mathbb{E}\big[\xi_{jt}(f_0, \beta_0)h_D^*(z_{jt})\big] \neq 0 \implies \forall q \in \mathbb{R}^+, \ \ \mathbb{P}(S_T(h_D^*, \mathcal{F}_0, \hat{\theta}) > q) \to 1$$

$\square$

Proof of Proposition 3.4.
Let $\mathcal{H}$ the set of measurable functions of z_{jt}, we want to show under $\bar{H}_a$:

$$\forall \alpha \in \mathbb{R}^*, \quad \alpha\mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}] \in \arg\max_{h \in \mathcal{H}} \operatorname{corr}(\xi_{jt}(f_0, \beta_0), h(z_{jt}))$$

We proceed in 2 steps. First, we derive the upper bound by showing that for any $h \in \mathcal{H}$, we have:

$$\operatorname{corr}\left(\xi_{jt}(f_0, \beta_0), h(z_{jt})\right) \leq \sqrt{\frac{\operatorname{var}\left(\mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]\right)}{\operatorname{var}(\xi_{jt}(f_0, \beta_0))}}$$

To do so, we use the definition of the conditional expectation and the Cauchy Schwarz inequality. First, notice that we have: $\mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}] = \mathbb{E}[\xi_{jt}(f_0, \beta_0)|z_{jt}]$. By definition of the conditional expectation, we have for any $h \in \mathcal{H}$,

$$\mathbb{E}[h(z_{jt})\xi_{jt}(f_0, \beta_0)] = \mathbb{E}[h(z_{jt})\mathbb{E}[\xi_{jt}(f_0, \beta_0)|z_{jt}]]$$

It follows that:

$$|\text{cov}\left(h(z_{jt}), \xi_{jt}(f_0, \beta_0)\right)| = \text{cov}\left(h(z_{jt}), \mathbb{E}[\xi_{jt}(f_0, \beta_0)|z_{jt}]\right) \leq \sqrt{\text{var}(h(z_{jt}))\text{var}\left(\mathbb{E}[\xi_{jt}(f_0, \beta_0)|z_{jt}]\right)}$$

The inequality comes from the Cauchy Schwarz inequality. The result follows by using the definition of the correlation coefficient.

Second, we show that the upper bound is reached by taking for any $\alpha \in \mathbb{R}^*$, $h^*(z_{jt}) = \alpha\mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]$.

$$\begin{aligned}
\text{cov}\left(\xi_{jt}(f_0, \beta_0), \alpha\mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]\right) &= \alpha \ \text{cov}\left(\Delta_{0,a}^{\xi_{jt}}, \mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]\right) \\
&= \alpha \ \text{var}\left(\mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]\right)
\end{aligned}$$

Consequently,

$$\text{corr}\left(\xi_{jt}(f_0, \beta_0), h^*(z_{jt})\right) = \frac{\alpha}{\sqrt{\alpha^2}}\sqrt{\frac{\text{var}\left(\mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]\right)}{\text{var}(\xi_{jt}(f_0, \beta_0))}} \implies |\text{corr}\left(\xi_{jt}(f_0, \beta_0), h^*(z_{jt})\right)| = \sqrt{\frac{\text{var}\left(\mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]\right)}{\text{var}(\xi_{jt}(f_0, \beta_0))}}$$

□

B.2.1 Connection with optimal instruments

In the parametric case, the BLP parameter θ is identified by the following non-linear conditional moment restriction $\mathbb{E}[\xi_{jt}(\theta)|z_{jt}] = 0$. The derivation of the optimal instruments in this context has been studied by Amemiya (1974). For an arbitrary choice of $h_E(z_{jt})$, the GMM estimator with the 2-step efficient weighting matrix has the following asymptotic distribution:

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}\left(0, (\Gamma(\mathcal{F}_0, \theta, h_E)'\Omega(\mathcal{F}_0, h_E)^{-1}\Gamma(\mathcal{F}_0, \theta, h_E))^{-1}\right)$$

with the same notations as previously:

$$\begin{aligned}
\Omega(\mathcal{F}_0, h_E) &= \mathbb{E}\left[\left(\sum_j \xi_{jt}(\theta)h_E(z_{jt})\right)\left(\sum_j h_E(z_{jt})\xi_{jt}(\theta)\right)'\right] \\
\Gamma(\mathcal{F}_0, \theta_0, h_E) &= \mathbb{E}\left[\sum_j h_E(z_{jt})\frac{\partial \xi_{jt}(\theta)}{\partial \tilde{\theta}'}\right]
\end{aligned}$$

For the sake of exposition, we will assume that unobserved demand shock ξ_{jt} is independent across observations, namely: $\mathbb{E}\left[\xi_{jt}(\theta)\xi_{j't}(\theta)|z_t\right] = 0$ for $j \neq j'$. The general case extends naturally. The optimal instrument $h_E^*(z_{jt})$ are chosen to minimize the asymptotic variance covariance matrix. We derive the form of the optimal instruments

in the context of BLP by adapting well known results in Chamberlain (1987) and Amemiya (1974)

Lemma 2.1 *Optimal instruments in the BLP model*

In our setting and assuming $f \in \mathcal{F}_0$, the optimal instruments $h_E^*(z_{jt})$ write:

$$h_E^*(z_{jt}) = \mathbb{E}[\xi_{jt}(\theta)^2|z_{jt}]^{-1}\mathbb{E}\left[\frac{\partial \xi_{jt}(\theta)}{\partial \theta}\middle| z_{jt}\right]$$

and the corresponding efficiency bound (obtained by setting $h_E = h_E^*$) writes:

$$V^* = \mathbb{E}\left[\sum_j \mathbb{E}\left[\frac{\partial \xi_{jt}(\theta)}{\partial \theta}\middle| z_{jt}\right]\mathbb{E}\left[\frac{\partial \xi_{jt}(\theta)}{\partial \theta}\middle| z_{jt}\right]' \mathbb{E}[\xi_{jt}(\theta)^2|z_{jt}]^{-1}\right]^{-1}$$

Proof. To shorten the notations, we denote: $\sigma^2(z_{jt}) = \mathbb{E}[\xi_{jt}(\theta)^2|z_{jt}]$ and $d(z_{jt}) = \mathbb{E}\left[\frac{\partial \xi_{jt}(\theta)}{\partial \theta}\middle| z_{jt}\right]$. Likewise, we define

$$\Omega_0(h_E) = \mathbb{E}\left[\sum_j \mathbb{E}[\xi_{jt}(\theta)^2|z_{jt}] h_E(z_{jt}) h_E(z_{jt})'\right]$$

We want to prove that for any set of instruments $h_E(z_{jt})$ that $V^*(z_{jt}) - \Gamma_0(h_E)'\Omega_0(h_E)^{-1}\Gamma_0(h_E)$ matrix is semi definite positive.

$$\begin{aligned}
&V^*(z_{jt}) - \Gamma_0(h_E)\Omega_0(h_E)^{-1}\Gamma_0(h_E)' = \\
&= \mathbb{E}\left[\sum_j d(z_{jt})d(z_{jt})'\sigma^2(z_{jt})\right] - \mathbb{E}\left[\sum_j \frac{\partial \xi_{jt}(\theta)}{\partial \theta} h_E(z_{jt})'\right]\Omega_0(h_E)^{-1}\mathbb{E}\left[\sum_j \frac{h_E(z_{jt})\partial \xi_{jt}(\theta)}{\partial \theta}'\right] \\
&= \mathbb{E}\left[\sum_j d(z_{jt})d(z_{jt})'\sigma^{-2}(z_{jt})\right] - \mathbb{E}\left[\sum_j d(z_{jt})h_E(z_{jt})'\right]\mathbb{E}\left[\sum_j \sigma^2(z_{jt})h_E(z_{jt})h_E(z_{jt})'\right]\mathbb{E}\left[\sum_j h_E(z_{jt})d(z_{jt})'\right] \\
&= \mathbb{E}\left[\mathbf{\tilde{D}(z_{jt})}'\mathbf{\tilde{D}(z_{jt})}\right] - \mathbb{E}\left[\mathbf{\tilde{D}(z_{jt})}'\mathbf{\tilde{H}_E(z_{jt})}\right]\mathbb{E}\left[\mathbf{\tilde{H}_E(z_{jt})}'\mathbf{\tilde{H}_E(z_{jt})}\right]^{-1}\mathbb{E}\left[\mathbf{\tilde{H}_E(z_{jt})}'\mathbf{\tilde{D}(z_{jt})}\right]
\end{aligned}$$

The second line comes from law of iterated expectations. Third line is a matricial way to rewrite the second line. $\mathbf{\tilde{D}(z_{jt})}$ a matrix which stacks $d(z_{jt})/\sigma(z_{jt})$ over the set of products (each line corresponds to one product j). Likewise, let $\tilde{\mathbf{H}}_{\mathbf{E}}(\mathbf{z_{jt}})$ a matrix which stacks $h_E(z_{jt})\sigma(z_{jt})$ over the set of products (each line corresponds to one product j). Now let us define the following matrices.

$$\tilde{\mathbf{X}} = \begin{pmatrix} \tilde{\mathbf{D}}(\mathbf{z_{jt}}) & \tilde{\mathbf{H}}_{\mathbf{E}}(\mathbf{z_{jt}}) \end{pmatrix} \text{ and } \tilde{\mathbf{M}} = \begin{pmatrix} \mathbf{I}_{|\theta_0|} & -\mathbb{E}\Big[\tilde{\mathbf{D}}(\mathbf{z_{jt}})'\tilde{\mathbf{H}}_{\mathbf{E}}(\mathbf{z_{jt}})\Big]\mathbb{E}\Big[\tilde{\mathbf{H}}_{\mathbf{E}}(\mathbf{z_{jt}})'\tilde{\mathbf{H}}_{\mathbf{E}}(\mathbf{z_{jt}})\Big]^{-1} \end{pmatrix}'$$

We have: $V^*(z_{jt}) - \Gamma_0(h_E)\Omega_0(h_E)^{-1}\Gamma_0(h_E) = \tilde{\mathbf{M}}'\mathbb{E}[\tilde{\mathbf{X}}'\tilde{\mathbf{X}}]\tilde{\mathbf{M}}$

The matrix above is clearly semi definite positive. □

B.3 Feasible most powerful instrument

B.3.1 Local approximation of the MPI

Proof of Proposition 4.1

Proof. First, we define $s_t^0 = \rho(\delta_t, x_{2t}, f_0(.|\lambda_0))$ with δ_t the true mean utility. From Lemma 2.2 ρ^{-1} is $\mathcal{C}^\infty$ and in particular, ρ^{-1} is $\mathcal{C}^1$. Thus, the Taylor expansion of $\rho^{-1}(s_t^0, x_{2t}, f_0(.|\lambda_0))$ around s_t writes:

$$\rho^{-1}(s_t^0, x_{2t}, f_0(.|\lambda_0)) = \rho^{-1}(s_t, x_{2t}, f_0(.|\lambda_0)) + \frac{\partial \rho^{-1}(s_t, x_{2t}, f_0(.|\lambda_0))}{\partial s}\bigg|_{s=s_t}(s_t^0 - s_t) + o\big(||s_t^0 - s_t||\big)$$
$$\delta_t = \rho^{-1}(s_t, x_{2t}, f_0(.|\lambda_0)) + \frac{\partial \rho^{-1}(s_t, x_{2t}, f_0(.|\lambda_0))}{\partial s}\bigg|_{s=s_t}(s_t^0 - s_t) + o\big(||s_t^0 - s_t||\big)$$

We now derive an expression for the first derivative of the inverse function. We make use of Lemma 2.3: for any $\delta \in \mathbb{R}^J$, $\frac{\partial \rho(\delta, x_{2t}, f)}{\partial \delta}$ is invertible.

$$\frac{\partial \rho(\rho^{-1}(s_t, x_{2t}, f_0(.|\lambda_0)), x_{2t}, f_0(.|\lambda_0))}{\partial s} = I_J \iff \frac{\partial \rho^{-1}(s_t, x_{2t}, f_0(.|\lambda_0))}{\partial s}\left(\frac{\partial \rho(\rho^{-1}(s_t, x_{2t}, f_0(.|\lambda_0)), x_{2t}, f_0(.|\lambda_0))}{\partial \rho^{-1}(s_t, x_{2t}, f_0(.|\lambda_0))}\right) = I_J$$
$$\iff \frac{\partial \rho^{-1}(s_t, x_{2t}, f_0(.|\lambda_0))}{\partial s} = \left(\frac{\partial \rho(\delta_t^0, x_{2t}, f_0(.|\lambda_0))}{\partial \delta}\right)^{-1}$$

with $\delta_t^0 = \rho^{-1}(s_t, x_{2t}, f_0(.|\lambda_0))$.Consequently,

$$\underbrace{\rho^{-1}(s_t, x_{2t}, f_0(.|\lambda_0)) - \delta_t}_{\Delta(s_t, x_{2t}, f_0, f_a)} = -\left(\frac{\partial \rho(\delta_t^0, x_{2t}, f_0(.|\lambda_0))}{\partial \delta}\right)^{-1} (s_t^0 - s_t) + o\big(||s_t^0 - s_t||\big) \tag{B.13}$$

with $\delta_t^0 = \rho_j^{-1}(s_t, x_{2t}, f_0(.|\lambda_0))$

Now let us show that there exists a constant M such that $||s_t^0 - s_t|| \leq M\tau(f_0(.|\lambda_0) - f_a)$. with $\tau(f_0 - f_a) = \int_{\mathbb{R}^{K_2}} |f_0(v|\lambda_0) - f_a(v)|dv$. Norms are equivalent in a finite vectorial space and without loss of generality, we will derive the results with the L_1 norm. By definition:

$$s_t^0 - s_t = \int_{\mathbb{R}^{K_2}} \frac{exp(\delta_t + x_{2t}v)}{1 + \sum_{k=1}^J \exp\left\{\delta_{kt} + x'_{2jk}v\right\}} (f_0(v|\lambda_0) - f_a(v))dv$$

Taking the L_1 norm of this vector:

$$\begin{aligned}
||s_t^0 - s_t||_1 &= \sum_{j=1}^J \left| \int_{\mathbb{R}^{K_2}} \frac{\exp(\delta_{jt} + x_{2jt}v)}{1 + \sum_{k=1}^J \exp\left\{\delta_{kt} + x'_{2jk}v\right\}} (f_0(v|\lambda_0) - f_a(v))dv \right| \\
&\leq \sum_{j=1}^J \int_{\mathbb{R}^{K_2}} \underbrace{\left| \frac{\exp(\delta_{jt} + x_{2jt}v)}{1 + \sum_{k=1}^J \exp\left\{\delta_{kt} + x'_{2jk}v\right\}} \right|}_{\leq 1} |f_0(v|\lambda_0) - f_a(v)|dv \\
&\leq J \int_{\mathbb{R}^{K_2}} |f_0(v|\lambda_0) - f_a(v)|dv = J\tau(f_0(.|\lambda_0) - f_a)
\end{aligned}$$

This proves the statement. As a consequence, we have: $||s_t^0 - s_t||_1 = O(\tau(f_0(.|\lambda_0) - f_a))$ and $o(||s_t^0 - s_t||) = o(\tau(f_0(.|\lambda_0) - f_a))$

The problem with the term $s_t^0 - s_t$ is that it is an expression of δ_t which we do not know under misspecification. As we want to be able to compute this approximation of the error term, it is not convenient in practice to have an expression which depends on δ_t. On the other hand, we know δ_t^0 and thus, the simple idea that we exploit is to take a Taylor expansion of the term above around δ_t^0. First, let us remark that from equation B.13, we have that:

$$||\delta_t - \delta_t^0|| = ||\delta_t - \rho^{-1}(s_t, x_{2t}, f_0(.|\lambda_0)|| = O(||s_t^0 - s_t||) = O(\tau(f_0(.|\lambda_0) - f_a))$$

Now let us take the Taylor expansion of $s_t^0 - s_t$ around δ_t^0:

$$s_t^0 - s_t = \int_{\mathbb{R}^{K_2}} \frac{exp(\delta_t^0 + x_{2t}v)}{1 + \sum_{k=1}^J \exp\left\{\tilde{\delta}_{kt} + x'_{2jk}v\right\}}(f_0(v|\lambda_0) - f_a(v))dv$$
$$+ \underbrace{\int_{\mathbb{R}^{K_2}} \frac{\partial}{\partial\delta'}\left\{\frac{exp(\delta_t^0 + x_{2t}v)}{1 + \sum_{k=1}^J \exp\left\{\delta_{kt}^0 + x'_{2jk}v\right\}}\right\}(\delta_t - \delta_t^0)(f_0(v|\lambda_0) - f_a(v))dv}_{B} + o\big(||\delta_t - \delta_t^0||\big)$$

From what precedes, we know that $o(||\delta_t - \delta_t^0||) = o(\tau(f_0(.|\lambda_0) - f_a))$. Now, let us show that term B in the previous expansion is also $o(\tau(f_0(.|\lambda_0) - f_a))$. Again taking the L_1 norm:

$$||B||_1 = \sum_{j=1}^J \left| \sum_{l=1}^J \int_{\mathbb{R}^{K_2}} \frac{\partial}{\partial\delta_l}\left\{\frac{\exp(\delta_{jt}^0 + x'_{2jt}v)}{1 + \sum_{k=1}^J \exp\left\{\tilde{\delta}_{kt} + x'_{2jk}v\right\}}\right\}(\delta_{lt} - \delta_{lt}^0)(f_0(v|\lambda_0) - f_a(v))dv \right|$$
$$\leq \sum_{j=1}^J \sum_{l=1}^J \int_{\mathbb{R}^{K_2}} \underbrace{\left|\frac{\partial}{\partial\delta_l}\left\{\frac{\exp(\delta_{jt}^0 + x'_{2jt}v)}{1 + \sum_{k=1}^J \exp\left\{\delta_{kt}^0 + x'_{2jk}v\right\}}\right\}\right|}_{\leq 1} |\delta_{lt} - \tilde{\delta}_{lt}||f_0(v|\lambda_0) - f_a(v)|dv$$
$$\leq J^2\tau(f_0(.|\lambda_0) - f)O(\tau(f_0(.|\lambda_0) - f_a)) = O(\tau(f_0(.|\lambda_0) - f_a)^2) = o(\tau(f_0(.|\lambda_0) - f_a))$$

Thus, $||B||_1 = o(\tau(f_0(.|\lambda_0) - f_a))$ and by combining all the results together, we get the final result. When $f_0(.|\lambda_0)$ gets ''close " to f_a, we have the following approximation:

$$\Delta(s_t, x_{2t}, f_0, f_a) = \left(\frac{\partial\rho(\delta_t^0, x_{2t}, f_0(.|\lambda_0))}{\partial\delta}\right)^{-1} \int_{\mathbb{R}^{K_2}} \frac{\exp(\delta_t^0 + x_{2t}v)}{1 + \sum_{k=1}^J \exp\left\{\delta_{kt}^0 + x'_{2jk}v\right\}}(f_a(v) - f_0(v|\lambda_0))dv$$
$$+ o(\tau(f_a - f_0(.|\lambda_0)))$$

$\delta_t^0 = \rho^{-1}(s_t, x_{2t}, f_0(.|\lambda_0))$ and $\tau(f_a - f_0(.|\lambda_0)) = \int_{\mathbb{R}^{K_2}} |f_a(v) - f_0(.|\lambda_0)(v)|dv$. □

B.3.2 Global approximation of the MPI

Derivation of $\Delta_j(s_t, x_{2t}, f_0, f_a)$

Proof.

$$1 = \frac{\rho(\delta_{jt}, x_{2t}, f_a)}{\rho(\delta_{jt}^0, x_{2t}, f_0)} = \frac{\int_{\mathbb{R}^{K_2}} \frac{\exp(\delta_{jt} + x'_{2jt}v)}{1+\sum_{k=1}^J \exp\left\{\delta_{kt} + x'_{2kt}v\right\}} f_a(v)dv}{\int_{\mathbb{R}^{K_2}} \frac{\exp(\delta_{jt}^0 + x'_{2jt}v)}{1+\sum_{k=1}^J \exp\left\{\delta_{kt}^0 + x'_{2kt}v\right\}} f_0(v)dv} \iff \frac{\exp(\delta_{jt}^0)}{\exp(\delta_{jt})} = \frac{\int_{\mathbb{R}^{K_2}} \frac{\exp(x_{2t}v)}{1+\sum_{k=1}^J \exp\left\{\delta_{kt} + x'_{2kt}v\right\}} f_a(v)dv}{\int_{\mathbb{R}^{K_2}} \frac{\exp(x'_{2jt}v)}{1+\sum_{k=1}^J \exp\left\{\delta_{jt}^0 + x'_{2kt}v\right\}} f_0(v)dv}$$

□

B.3.3 Approximation of the MPI in the mixed logit case

Proof of Proposition 1.1. By definition, we have:

$$\begin{aligned} g_j(x_i, ., f) :\ & \mathbb{R}^{K_1} \to [0,1] \\ & \beta \mapsto \int_{\mathbb{R}^{K_2}} \frac{\exp\left\{x'_{ij1}\beta + x'_{2ij}v\right\}}{1+\sum_{k=1}^{J}\exp\left\{x'_{ik1}\beta + x'_{2ik}v\right\}} f(v)dv \end{aligned}$$

g is $\mathcal{C}^\infty$ on $\mathbb{R}^{K_1}$. Thus, we can take a first order Taylor expansion of $g_j(x_i, ., f_1)$ around β_0^*:

$$g_j(x_i, \beta_1, f_1) = g_j(x_i, \beta_0^*, f_1) + \left.\frac{\partial g(x_i, \beta, f_1)}{\partial \beta}\right|_{\beta=\beta_0^*} (\beta_1 - \beta_0^*) + o(||\beta_1 - \beta_0^*||)$$

This yields immediately,

$$g(x_i, \beta_0^*, f_0^*) - g(x_i, \beta_1, f_1) = \int_{\mathbb{R}^{K_2}} \frac{exp(x'_{1ij}\beta_0^* + x'_{2ij}v)}{1+\sum_{k=1}^{J}\exp\left\{x'_{1ik}\beta_0^* + x'_{2ik}v\right\}} (f_0^*(v) - f_1(v))dv +$$
$$\left.\frac{\partial g(x_i, \beta, f_1)}{\partial \beta}\right|_{\beta=\beta_0} (\beta_a - \beta_0) + o(||\beta_- \beta_0^*||)$$

Now let us show that $||\beta_1 - \beta_0^*|| = ...$

By construction, the pseudo true values β_0^* and β_1^* maximize the conditional expectation of the log-likelihood:

$$\beta_0^* = \underset{\beta\in\mathbb{R}^{K_1}}{argmax}\, \mathbb{E}\big[L(x_i, y_i, \beta, f_0^*)\big|x_i\big] \text{ with } L(x_i, y_i, \beta, f_0^*) = \sum_{j=0}^{J} \mathbf{1}\{y_{ij} = 1\} log(g_j(x_i, \beta, f_0^*))$$

The same goes for β_1^*:

$$\beta_1^* = \underset{\beta\in\mathbb{R}^{K_1}}{argmax}\, \mathbb{E}\big[L(x_i, y_i, \beta, f_1^*)\big|x_i\big] \text{ with } L(x_i, y_i, \beta, f_1^*) = \sum_{j=0}^{J} \mathbf{1}\{y_{ij} = 1\} log(g_j(x_i, \beta, f_1^*))$$

When the true distribution of RC is f_1, we have:

$$\mathbb{E}\big[L(x_i, y_i, \beta, f_0^*)\big|x_i\big] = \sum_{j=0}^{J} g_j(x_i, \beta_1^*, f_1) log(g_j(x_i, \beta, f_0^*))$$
$$\mathbb{E}\big[L(x_i, y_i, \beta, f_1)\big|x_i\big] = \sum_{j=0}^{J} g_j(x_i, \beta_1^*, f_1) log(g_j(x_i, \beta, f_1))$$

$\square$

B.4 Specification test: composite hypothesis

In this section, we prove Theorem 5.1, which is the main asymptotic result of the paper. The section is organized as follows. First, we establish the equivalence between the moment condition around which we build our test $\mathbb{E}\left[\sum_{jt} \xi_{jt}(f_0(\cdot|\lambda_0), \beta_0) h_D(z_{jt})\right] = 0$ and the one characterizing $H_0' : \mathbb{E}\left[\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0) h_D(z_{jt})\right] = 0$. Then, we introduce the notations used in the proofs and we decompose $\widehat{\xi}$ according to the BLP approximations. Second, we provide technical lemmas which prove that under the assumptions in E, the BLP approximations vanish asymptotically. Third, we prove that the BLP estimator is consistent and asymptotically normal. Finally, we prove the main theorem and we show that under the null the test is pivotal in the 2 polar cases described in the main text.

B.4.1 Equivalence between moment conditions

Let $h_D(z_{jt})$ our detection instruments. For conciseness, we omit the dependence in f_0 and denote $\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0) = \xi_{jt}(\theta_0)$. We want to prove that the following two moment conditions are equivalent:

$$\mathbb{E}\left[\xi_{jt}(\theta_0) h_D(z_{jt})\right] = 0 \iff \mathbb{E}\left[\sum_{j=1}^{J} \xi_{jt}(\theta_0) h_D(z_{jt})\right] = 0$$

Let R_t a categorial random variable which exogenously selects a product j with probability $\frac{1}{J}$. Formally, we have $(\xi_{jt}(\theta_0), z_{jt}) \perp R_{jt}$. By construction, we have:

$$\begin{aligned}\mathbb{E}\left[\xi_{jt}(\theta_0) h_D(z_{jt})\right] &= \sum_{k=1}^{J} \mathbb{E}\left[\xi_{kt}(\theta_0) h_D(z_{kt}) R_{kt}\right] = \sum_{k=1}^{J} \mathbb{E}\left[\xi_{kt}(\theta_0) h_D(z_{kt})\right] \mathbb{E}[R_{kt}] \\ &= \frac{1}{J} \mathbb{E}\left[\sum_{k=1}^{J} \xi_{kt}(\theta_0) h_D(z_{kt})\right]\end{aligned}$$

The second line results from independence of $(\xi_{jt}(\theta_0), z_{jt})$ and R_{jt}. This proves the result.

B.4.2 Notations

In the proofs, we will adopt the following notations. If the derivations are done under the parametric assumption $H_0 : f \in \mathcal{F}_0$ then we omit the dependence in f_0 and interchangeably use $\xi_{jt}(f_0(.|\lambda),\beta)$ and $\xi_{jt}(\theta)$. We also omit the dependence of the BLP pseudo true value in W and $h_E(z_{jt})$ [37]. Then define the following objectives of the GMM minimization

$$\hat{\mathcal{Q}}_T(\tilde{\theta}) = \left(\frac{1}{T}\sum_{j,t}\widehat{\xi}_{jt}(\tilde{\theta})h_E(z_{jt})\right)'\hat{W}\left(\frac{1}{T}\sum_{j,t}\widehat{\xi}_{jt}(\tilde{\theta})h_E(z_{jt})\right)$$
$$\mathcal{Q}_T(\tilde{\theta}) = \left(\frac{1}{T}\sum_{j,t}\xi_{jt}(\tilde{\theta})h_E(z_{jt})\right)'\hat{W}\left(\frac{1}{T}\sum_{j,t}\xi_{jt}(\tilde{\theta})h_E(z_{jt})\right)$$
$$\mathcal{Q}(\tilde{\theta}) = \mathbb{E}\left[\sum_{j}\xi_{jt}(\tilde{\theta})h_E(z_{jt})\right]'W\mathbb{E}\left[\sum_{j}\xi_{jt}(\tilde{\theta})h_E(z_{jt})\right]$$

We also define the following moments

$$\hat{g}_T(\tilde{\theta},h) = \frac{1}{T}\sum_{jt}\hat{\xi}_{jt}(\tilde{\theta})h(z_{jt})$$
$$g_T(\tilde{\theta},h) = \frac{1}{T}\sum_{jt}\xi_{jt}(\tilde{\theta})h(z_{jt})$$
$$g(\tilde{\theta},h) = \mathbb{E}\left[\sum_{j}\xi_{jt}(\tilde{\theta})h(z_{jt})\right]$$

And recall the definition of $\Gamma(\mathcal{F}_0,\tilde{\theta},h)$ which is used interchangeably with $\Gamma(\tilde{\theta},h)$

$$\hat{\Gamma}_T(\tilde{\theta},h) = \frac{1}{T}\sum_{j,t}h(z_{jt})\frac{\partial}{\partial\theta}\widehat{\xi}_{jt}(\tilde{\theta})'$$
$$\Gamma_T(\tilde{\theta},h) = \frac{1}{T}\sum_{j,t}h(z_{jt})\frac{\partial}{\partial\theta}\xi_{jt}(\tilde{\theta})'$$
$$\Gamma(\tilde{\theta},h) = \mathbb{E}\left[\sum_{j}h(z_{jt})\frac{\partial}{\partial\theta}\xi_{jt}(\tilde{\theta})'\right]$$

Furthermore, unless specified, all limits are taken with respect to T; Additionally, we denote by the expression $X = o_P(T^\kappa)$ a random variable or statistic which is asymptotically degenerate of order T^a, ie $X = o_P(T^\kappa) \Leftrightarrow \forall e>0\, \mathbb{P}(|X|T^{-\kappa}>e) \underset{T\to\infty}{\to} 0$, and denote by $X = O_p(T^\kappa)$ a random variable which is (bounded in probability) of order T^κ, ie $\forall e_1>0 \exists e_2>0, \exists T_N : \forall T \geqslant T_N\, \mathbb{P}(|X|T^{-\kappa}>e_2)<e_1$. Properties of $o_P(1)$ and $O_P(1)$ random variables are used throughout these proofs.

[37] The BLP pseudo true value depends on W and $h_E(z_{jt})$ when the model is misspecified

B.4.3 Feasible Structural Error and BLP approximations

We now decompose the difference between the true structural error $\xi_{jt}(\tilde{\theta})$ and the feasible structural error $\hat{\xi}_{jt}(\tilde{\theta})$ in terms of the different approximations involved in the derivation of the feasible structural error $\hat{\xi}_{jt}(\tilde{\theta})$. In market t given an assumption $\mathcal{F}_0$, a parameter $\tilde{\lambda}$, market shares s_t and product characteristics with preference heterogeneity x_{2t} there exists a unique $\delta_t \in \mathbb{R}^J$ such that $s_t = \rho(\delta_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))$ (Brouwer's fixed point theorem, see Berry (1994)) so that $\delta_t = \rho^{-1}(s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))$. There is no closed form for $\rho^{-1}(s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))$ so the NFP algorithm is used. Denote as C the contraction used to find the mean utilities which solve the demand equal market share constraint

$$C(\cdot, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) : \delta \in \mathbb{R}^J \mapsto \delta + log(s_t) - log(\rho(\delta, x_{2t}, f_0(\cdot|\tilde{\lambda})))$$

So that for some starting mean utility $\delta_0 \in \mathcal{B} \subset \mathbb{R}^J$ where $\mathcal{B}$ is bounded, the mean utility obtained via NFP at the limit is equal to the unique vector which solves the constraint

$$\delta_t(f_0(\cdot|\tilde{\lambda})) = \rho^{-1}(s_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) = \underset{H\to\infty}{lim}\ C^{(H)}(\delta_0, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))$$

Similarly the error generated by $(f_0(\cdot|\tilde{\lambda}, \tilde{\beta}))$ can be obtained from NFP at the limit

$$\xi_t(f_0(\cdot|\tilde{\lambda}), \tilde{\beta}) = \delta_t(f_0(\cdot|\tilde{\lambda})) - x_{1t}\tilde{\beta} = \underset{H\to\infty}{lim}\ C^{(H)}(\delta_0, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) - x_{1t}\tilde{\beta}$$

This way we obtain a vector of mean utilities for each market t. There are 3 approximations to consider, market shares are not truly observed, the demand integral has to be simulated, and the contraction is never taken to its limit, so define $\widehat{\xi}(f_0, \tilde{\lambda}))$ $\widehat{\delta}(f_0, \tilde{\lambda}))$ and $\hat{C}$ for some starting value δ_0

$$\widehat{\xi}_t(f_0(\cdot|\tilde{\lambda}), \tilde{\beta}) = \hat{C}^{(H)}(\delta_0, \hat{s}_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) - x_{1jt}\tilde{\beta}, \qquad \widehat{\delta}(f_0, \tilde{\lambda})) = \hat{C}^{(H)}(\delta_0, \hat{s}_t, x_{2t}, f_0, \tilde{\lambda}))$$
$$\hat{C} : \delta \mapsto \delta + log(\hat{s}_t) - log(\hat{\rho}(\delta, x_{2t}, f_0(\cdot|\lambda_0)))$$

Consequently, we decompose the difference between the error generated by $(f_0(\cdot|\tilde{\lambda}), \tilde{\beta})$ and its feasible approximation into 3 differences

$$\begin{aligned}
\xi_{jt}(f_0(\cdot|\tilde{\lambda}), \tilde{\beta}) - \widehat{\xi}_{jt}(f_0(\cdot|\tilde{\lambda}), \tilde{\beta}) &= \delta_{jt}(f_0(\cdot|\tilde{\lambda})) - \widehat{\delta}_{jt}(f_0(\cdot|\tilde{\lambda})) \\
&= \underset{H\to\infty}{lim}\ C_j^{(H)}(\delta_0, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))) - \hat{C}_j^{(H)}(\delta_0, \hat{s}_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) \\
&= \underset{H\to\infty}{lim}\ C_j^{(H)}(\delta_0, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))) - C_j^{(H)}(\delta_0, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) \\
&\quad + C_j^{(H)}(\delta_0, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))) - C_j^{(H)}(\delta_0, \hat{s}_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) \\
&\quad + C_j^{(H)}(\delta_0, \hat{s}_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) - \hat{C}_j^{(H)}(\delta_0, \hat{s}_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) \\
&\equiv \rho_j^{-1}(s_t, x_{2t}, f_0(\cdot|\tilde{\lambda})) - D_j(\rho, s_t, \tilde{\lambda}) \\
&\quad + D_j(\rho, s_t, \tilde{\lambda}) - D_j(\rho, \hat{s}_t, \tilde{\theta}) \\
&\quad + D_j(\rho, \hat{s}_t, \tilde{\theta}) - D_j(\hat{\rho}, \hat{s}_t, \tilde{\theta})
\end{aligned}$$

In the fourth line, we simply introduce shortened notations for the same objects.

B.4.4 Technical lemmas

The 1st and 2nd lemma establish the smoothness of ρ^{-1} and the invertibility of the Jacobian matrix of ρ with respect to δ. In the 3rd lemma, we derive the Lipschitz constant of the contraction and we prove that it is bounded away from 0 and 1. The 4th lemma ensures that for key moments and quantities the BLP approximations can be ignored uniformly asymptotically.

Lemma 2.2 *ρ^{-1} is $\mathcal{C}^\infty$*

Proof. We know that the demand function ρ is $\mathcal{C}^\infty$ and invertible on $\mathbb{R}^J$. Moreover,$\forall \delta \in \mathbb{R}^J$, $\frac{\partial \rho(\delta, x_{2t}, f)}{\partial \delta} \neq 0$. As a consequence, $\rho^{-1} : [0,1]^J \to \mathbb{R}^J$ the inverse demand function is also $\mathcal{C}^\infty$. □

Lemma 2.3 *For any $\delta \in \mathbb{R}^J$, $\frac{\partial \rho(\delta, x_{2t}, f)}{\partial \delta}$ is invertible.*

Proof. $\frac{\partial \rho}{\partial \delta}$ is a $J \times J$ matrix such that $\left(\frac{\partial \rho}{\partial \delta}\right)_{j,k}$ is:

$$\frac{\partial \rho_j \left(\delta_t, x_{2t}, f\right)}{\partial \delta_{kt}} = \begin{cases} \int \mathcal{T}_{jt}(v) \left(1 - \mathcal{T}_{kt}(v)\right) f(v) dv & if \quad j = k \\ -\int \mathcal{T}_{jt}(v) \mathcal{T}_{kt}(v) f(v) dv & if \quad j \neq k \end{cases}$$

with $\mathcal{T}_{jt}(v) \equiv \frac{exp\{\delta_{jt} + x'_{2jt} v\}}{1 + \sum_{j'=1}^{J} exp\{\delta_{j't} + x'_{2j't} v\}}$

One can easily check that $\frac{\partial \rho}{\partial \delta}$ is strictly diagonally dominant. Indeed for each row j:

$$\left|\frac{\partial \rho_j \left(\delta_t, x_{2t}, f\right)}{\partial \delta_{kt}}\right| - \sum_{k \neq j} \left|\frac{\partial \rho_j \left(\delta_t, x_{2t}, f\right)}{\partial \delta_{kt}}\right| = \int \mathcal{T}_{jt}(v) \underbrace{\left(1 - \sum_{k=1}^{J} \mathcal{T}_{kt}(v)\right)}_{>0} f(v) dv > 0$$

□

Lemma 2.4 (Contraction mapping Lipschitz constant)
Given parametric assumption $\mathcal{F}_0$, under assumptions B-E, assume that starting mean utility δ_0 is in $\mathcal{B}$ where $\mathcal{B}$ is compact, then without loss of generality there exists some

$(\underline{a}, \bar{a}) \in \mathbb{R}^2$ *with* $\bar{a} > \underline{a}$ *such that for any* $b \in \mathcal{B}$ *for any* $j = 1, \ldots, J$ $\underline{a} \leqslant b_j \leqslant \bar{a}$, *furthermore denote by* $\mathcal{X}$ *the compact support of* x_{2jt}. *Then on* $\mathcal{B}$ *the map* $C(\cdot, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}_0))$ *is a contraction with Lipschitz constant*

$$\epsilon = \max_{j=1,\ldots,J} \sup_{a \in \mathcal{B}, b \in [0; \bar{a}-\underline{a}]^J, x_2 \in \mathcal{X}, \tilde{\lambda} \in \Lambda_0} 1 - \frac{\int \frac{exp(a_j+b_j+x'_{2j}v)}{\left(1+\sum_k \exp(a_k+b_k+x'_{2k}v)\right)^2} f_0(v|\tilde{\lambda})dv}{\int \frac{exp(a_j+b_j+x'_{2j}v)}{1+\sum_k \exp(a_k+b_k+x'_{2k}v)} f_0(v|\tilde{\lambda})dv}$$

which is in $(0; 1)$

Proof. This proof is inspired by the proof of the Theorem in Appendix 1 of Berry et al. (1995). Let $C_j(\cdot) \equiv C(\cdot, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}_0))$, we first determine the partial derivative of $C_j(\cdot)$

$$\begin{aligned}
\frac{\partial C_j(a)}{\partial a_j} &= 1 - \frac{1}{\rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))} \int \frac{exp(a_j + x'_{2kt}v)(1 + \sum_{k=1}^J exp(a_k + x'_{2kt}v)) - exp(2(a_j + x'_{2kt}v))}{(1 + \sum_{k=1}^J exp(a_k + x'_{2kt}v))^2} f_0(v|\tilde{\lambda})dv \\
&= \frac{1}{\rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))} \int \frac{exp(2(a_j + x'_{2jt}v))}{(1 + \sum_{k=1}^J exp(a_k + x'_{2kt}v))^2} f_0(v|\tilde{\lambda})dv \\
\frac{\partial C_j(a)}{\partial a_{j'}} &= \frac{1}{\rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))} \int \frac{exp(a_j + x'_{2jt}v) exp(a_{j'} + x'_{2j't}v)}{(1 + \sum_{k=1}^J exp(a_k + x'_{2kt}v))^2} f_0(v|\tilde{\lambda})dv
\end{aligned}$$

Note that for any $j = 1, \ldots, J$ all partial derivatives of $C_j(\cdot)$ are strictly positive and that the sum of its derivatives evaluated in a equals

$$\begin{aligned}
\sum_{k=1}^J \frac{\partial C_j(a)}{\partial a_k} &= \frac{1}{\rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))} \int \frac{exp(a_j + x'_{2jt}v) \sum_{k=1}^J exp(a_k + x'_{2kt}v)}{(1 + \sum_{k=1}^J exp(a_k + x'_{2kt}v))^2} f_0(v|\tilde{\lambda})dv \\
&= \frac{1}{\rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))} \int \frac{exp(a_j + x'_{2jt}v)(1 + \sum_{k=1}^J exp(a_k + x'_{2kt}v) - 1)}{(1 + \sum_{k=1}^J exp(a_k + x'_{2kt}v))^2} f_0(v|\tilde{\lambda})dv \\
&= 1 - \frac{\int \frac{exp(a_j + x'_{2jt}v)}{(1+\sum_{k=1}^J exp(a_k + x'_{2kt}v))^2} f_0(v|\tilde{\lambda})dv}{\rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))}
\end{aligned}$$

For any $(a_1, a_2) \in \mathcal{B}^2$ let $\tilde{a} = (||a_1 - a_2||_\infty, \ \dots \ , ||a_1 - a_2||_\infty) \in \mathbb{R}^J$ then

$$\begin{aligned}
C_j(a_1) - C_j(a_2) &= C_j(a_2 + a_1 - a_2) - C_j(a_2) \leqslant C_j(a_2 + \tilde{a}) - C_j(a_2) \\
&\leqslant \int_{0^J}^{||a_1 - a_2||_\infty^J} \frac{\partial C_j(a_2 + b)}{\partial a} db \\
&\leqslant ||a_1 - a_2||_\infty \sup_{a \in \mathcal{B}, b \in [0; \bar{a} - \underline{a}]^J} \sum_{k=1}^{J} \frac{\partial C_j(a + b)}{\partial a_k} \\
&\leqslant ||a_1 - a_2||_2 \max_{j=1,..J} \sup_{a \in \mathcal{B}, b \in [0; \bar{a} - \underline{a}]^J, x_2 \in \mathcal{X}, \tilde{\lambda} \in \Lambda_0} \sum_{k=1}^{J} \frac{\partial C_j(a + b)}{\partial a_k} \\
&\equiv ||a_1 - a_2||_2 \ \epsilon
\end{aligned}$$

where the 1st inequality holds because $C_j(\cdot)$ is increasing in all its inputs, the 2nd inequality holds by the fundamental theorem of calculus and by the total derivative formula, the 3rd and 4th inequalities hold by properties of norms.

We now prove that $\sup_{a \in \mathcal{B}, b \in [0; \bar{a} - \underline{a}]^J, \tilde{\lambda} \in \Lambda_0} \sum_{k=1}^{J} \frac{\partial C_j(a+b)}{\partial a_k} \in (0; 1)$ which will imply that $\epsilon \in (0; 1)$. To do so we have to prove that $\sum_{k=1}^{J} \frac{\partial C_j(a, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))}{\partial a_k}$ is continuous in $(a, x_{2t}, \tilde{\lambda})$ and takes values in $(0; 1)$ almost surely, this way because $\mathcal{B}$, $\mathcal{X}$ and Λ_0 are compact by Weierstrass' extreme value Theorem the sum of partial derivatives will also take values in a compact which is inside $(0; 1)$, then the supremum will become a maximum which can be attained and which is inside $(0; 1)$. The sum of partial derivatives is almost surely in $(0; 1)$ because

$$
\begin{aligned}
&\int \frac{exp(a_j + x'_{2jt}v)}{(1+\sum_{k=1}^J exp(a_k + x'_{2kt}v))^2} f_0(v|\tilde{\lambda})dv - \rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))\\
&= \int \frac{exp(a_j + x'_{2jt}v)}{(1+\sum_{k=1}^J exp(a_k + x'_{2kt}v))^2} f_0(v|\tilde{\lambda})dv - \int \frac{exp(a_j + x'_{2jt}v)}{1+\sum_{k=1}^J exp(a_k + x'_{2kt}v)} f_0(v|\tilde{\lambda})dv\\
&= -\int \frac{exp(a_j + x'_{2jt}v)\sum_{k=1}^J exp(a_k + x'_{2kt}v)}{(1+\sum_{k=1}^J exp(a_k + x'_{2kt}v))^2} f_0(v|\tilde{\lambda})dv < 0\\
&\Rightarrow \frac{\int \frac{exp(a_j + x'_{2jt}v)}{1+\sum_{k=1}^J exp(a_k + x'_{2kt}v)} f_0(v|\tilde{\lambda})dv}{\rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))} < 1\\
&\Rightarrow \sum_{k=1}^J \frac{\partial C_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))}{\partial a_k} = 1 - \frac{\int \frac{exp(a_j + x'_{2jt}v)}{1+\sum_{k=1}^J exp(a_k + x'_{2kt}v)} f_0(v|\tilde{\lambda})dv}{\rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))} > 0\\
&\quad -\frac{\int \frac{exp(a_j + x'_{2jt}v)}{1+\sum_{k=1}^J exp(a_k + x'_{2kt}v)} f_0(v|\tilde{\lambda})dv}{\rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))} < 0\\
&\Rightarrow \sum_{k=1}^J \frac{\partial C_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))}{\partial a_k} = 1 - \frac{\int \frac{exp(a_j + x'_{2jt}v)}{1+\sum_{k=1}^J exp(a_k + x'_{2kt}v)} f_0(v|\tilde{\lambda})dv}{\rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))} < 1
\end{aligned}
$$

Continuity of the sum of the partial derivatives in (a, x_{2t}) is trivial, continuity in $\tilde{\lambda}$ also holds because $f_0(\cdot|\tilde{\lambda})$ must be continuously differentiable via Assumption D. $\forall e_1 > 0, \exists e_2 : \forall(\lambda_1, \lambda_2) : ||\lambda_1 - \lambda_2||_2 \leqslant e_2$ implies $|f_0(v|\lambda_1) - f_0(v|\lambda_2)| < e_1$ for all v which in turn implies

$$
\begin{aligned}
\forall x_2 \in \mathcal{X}, \forall a \in \mathcal{B} \quad & \left| \int \frac{exp(a_j + x'_{2j}v)}{1+\sum_{k=1}^J exp(a_k + x'_{2k}v)} (f_0(v|\lambda_1) - f_0(v|\lambda_2))dv \right|\\
& \leqslant \int \frac{exp(a_j + x'_{2j}v)}{1+\sum_{k=1}^J exp(a_k + x'_{2k}v)} |f_0(v|\lambda_1) - f_0(v|\lambda_2)|dv \leqslant e_1\\
& \left| \int \frac{exp(a_j + x'_{2j}v)}{(1+\sum_{k=1}^J exp(a_k + x'_{2k}v))^2} (f_0(v|\lambda_1) - f_0(v|\lambda_2))dv \right| \leqslant e_1
\end{aligned}
$$

thus both $\tilde{\lambda} \mapsto \rho_j(a, x_{2t}, f_0(\cdot|\tilde{\lambda}))$ and $\tilde{\lambda} \mapsto \int \frac{exp(a_j + x'_{2jt}v}{(1+\sum_{k=1}^J exp(a_k + x'_{2kt}v))^2} f_0(v|\tilde{\lambda})dv$ are continuous and so is their ratio. □

Lemma 2.5 (Uniform convergence of objective function wrt BLP approximations)

Given parametric assumption $\mathcal{F}_0$, under assumptions B-E and $\forall h$ which satisfies D

$$\underset{\tilde{\theta}\in\Theta_0}{sup}\ \sqrt{T}||\hat{g}_T(\tilde{\theta},h)-g_T(\tilde{\theta},h)||_2 \xrightarrow{P} 0$$
$$\underset{\tilde{\theta}\in\Theta_0}{sup}||\hat{\Gamma}_T(\tilde{\theta},h)-\Gamma_T(\tilde{\theta},h)||_2 \xrightarrow{P} 0$$
$$\underset{\tilde{\theta}\in\Theta_0}{sup}|\hat{\mathcal{Q}}_T(\tilde{\theta})-\mathcal{Q}(\tilde{\theta})| \xrightarrow{P} 0$$

Proof. Parts of this proof are inspired from Freyberger (2015). We prove the 3 statements of the lemma in order

1. Using the properties of the *sup*, the fact that $\forall(A,B)$ rv, $\forall e>0$, $\forall\alpha\in(0,1)$, $\mathbb{P}(A+B>e)\leqslant\mathbb{P}(A>\alpha e)+\mathbb{P}(B>(1-\alpha)e)$ and the previous decomposition of the difference between ξ and $\widehat{\xi}$ we can find an upper bound on the probability that that the difference between $\hat{g}_T(\cdot)$ and $g_T(\cdot)$ is above a deviation: For any $e_1>0$

$$\begin{aligned}\mathbb{P}(\underset{\tilde{\theta}}{sup}\ \sqrt{T}||\hat{g}_T(\theta,h)-g_T(\theta,h)||_2>e_1) &= \mathbb{P}(\underset{\tilde{\theta}}{sup}\ \sqrt{T}\frac{1}{T}||\sum_{j,t}(\widehat{\xi}_t(f_0(\cdot|\tilde{\lambda}),\tilde{\beta})-\xi_t(f_0(\cdot|\tilde{\lambda}),\tilde{\beta}))h(z_{jt})||_2>e_1) \\ &\leqslant \mathbb{P}(\underset{\tilde{\lambda}}{sup}\ \sqrt{T}||\frac{1}{T}\sum_{j,t}(\rho^{-1}(s_t,x_{2t},f_0(\cdot|\tilde{\lambda}_0))-D_j(\rho,s_t,\tilde{\lambda}))h(z_{jt})||_2>\frac{e_1}{3}) \\ &\quad+\mathbb{P}(\underset{\tilde{\lambda}}{sup}\ \sqrt{T}||\frac{1}{T}\sum_{j,t}(D_j(\rho,s_t,\tilde{\lambda})-D_j(\rho,\hat{s}_t,\tilde{\lambda}))h(z_{jt})||_2>\frac{e_1}{3}) \\ &\quad+\mathbb{P}(\underset{\tilde{\lambda}}{sup}\ \sqrt{T}||\frac{1}{T}\sum_{j,t}(D_j(\rho,\hat{s}_t,\tilde{\lambda})-D_j(\hat{\rho},\hat{s}_t,\tilde{\lambda}))h(z_{jt})||_2>\frac{e_1}{3})\end{aligned}$$

Then we can prove that each element of the upper bound converges to 0

(a) By properties of contractions and using Lemma 2.4 we have

$$|\rho^{-1}(s_t,x_{2t},f_0(\cdot|\tilde{\lambda}_0))-D_j(\rho,s_t,\tilde{\lambda})|\leqslant\epsilon^H|\rho^{-1}(s_t,x_{2t},f_0(\cdot|\tilde{\lambda}_0))-\delta_0|\leqslant\epsilon^H\kappa$$

for some constant κ which exists due to the compactness of Λ_0, $\mathcal{X}$ and $\mathcal{B}$. Thus using the iid nature of the data **??**(i), the speed of the NFP algorithm Assumption E(iii), the triangle inequality, Markov inequality and Cauchy-

Schwarz inequality the 1st element converges to 0

$$\begin{aligned}
&\mathbb{P}(\sup_{\tilde{\lambda}} \sqrt{T}||\frac{1}{T}\sum_{j,t}(\rho^{-1}(s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}_0)) - D_j(\rho, s_t, \tilde{\lambda}))h(z_{jt})||_2 > \frac{e_1}{3})\\
&\leqslant \mathbb{P}(\sqrt{T}\epsilon^H \kappa||\frac{1}{T}\sum_{j,t} h(z_{jt})||_2 > \frac{e_1}{3}) \leqslant \mathbb{P}(\sqrt{T}\epsilon^H \frac{1}{T}\sum_{j,t}||h(z_{jt})||_2 > \frac{e_1}{3})\\
&\leqslant \frac{3\kappa}{e_1}\sqrt{T}\epsilon^H \sum_j \sqrt{\mathbb{E}(||h(z_{jt})||_2^2)} \underset{T\to\infty}{\to} 0
\end{aligned}$$

(b) Note that D_j is continuously differentiable in $s \in (0;1)$ so that it is uniformly continuous in s. Indeed C is $\mathcal{C}^\infty$ in s so that

$$\frac{\partial D(\rho, s_t, \tilde{\lambda})}{\partial s} = \prod_{h=1}^{H} \frac{\partial C(C^{(h-1)}(\delta_0, s_t, x_{2t}, f_0(\cdot|\tilde{\lambda})), s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))}{\partial s}$$

Next because Λ_0 is compact it can be covered by some finite union of closed balls in $\mathbb{R}^{K_2}$, ie $\Lambda_0 \subset \cup_{c=1}^{N}\Lambda_{0,c}^N$ with $\forall c = 1, \ldots, N \quad \Lambda_{0,c}^N = \{\tilde{\lambda} : ||\tilde{\lambda} - \lambda_c||_2 \leqslant r_N\}$, $\lambda_c \in \Lambda_0$ and $r_N \underset{N\to\infty}{\to} 0$. Consequently

$$\begin{aligned}
&\mathbb{P}(\sup_{\tilde{\lambda}} \frac{1}{\sqrt{T}}||\sum_{j,t}(D_j(\rho, s_t, \tilde{\lambda}) - D_j(\rho, \hat{s}_t, \tilde{\lambda}))h_E(z_{jt})||_2 > \frac{e_1}{3})\\
&\leqslant \mathbb{P}(\max_{c=1,\ldots,N} \sup_{\tilde{\lambda}\in\Lambda_{0,c}^N} \frac{1}{\sqrt{T}}||\sum_{j,t}(D_j(\rho, s_t, \tilde{\theta}) - D_j(\rho, \hat{s}_t, \tilde{\theta}))h_E(z_{jt})||_2 > \frac{e_1}{3})\\
&\leqslant \sum_{c=1}^{N}\mathbb{P}(\sup_{\tilde{\lambda}\in\Lambda_{0,c}^N} \frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho, s_t, \tilde{\lambda}) - D_j(\rho, \hat{s}_t, \tilde{\lambda})|\ ||h_E(z_{jt})||_2 > \frac{e_1}{3})\\
&\leqslant \sum_{c=1}^{N}\mathbb{P}(\frac{1}{\sqrt{T}}||\sum_{j,t}(D_j(\rho, s_t, \lambda_c) - D_j(\rho, \hat{s}_t, \lambda_c))h_E(z_{jt})||_2 > \frac{e_1}{9})\\
&\quad + \sum_{c=1}^{N}\mathbb{P}(\sup_{\tilde{\lambda}\in\Lambda_{0,c}^N} \frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho, s_t, \tilde{\lambda}) - D_j(\rho, s_t, \lambda_c)|\ ||h_E(z_{jt})||_2 > \frac{e_1}{9})\\
&\quad + \sum_{c=1}^{N}\mathbb{P}(\sup_{\tilde{\lambda}\in\Lambda_{0,c}^N} \frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho, \hat{s}_t, \lambda_c) - D_j(\rho, \hat{s}_t, \tilde{\lambda})|\ ||h_E(z_{jt})||_2 > \frac{e_1}{9})
\end{aligned}$$

where the last inequality was obtained using the triangle inequality. Then by uniform continuity of D_j in s it follows that $\exists e_2 > 0$ such that $\forall c$ $\frac{1}{\sqrt{T}}||\sum_{j,t}(D_j(\rho, s_t, \lambda_c) - D_j(\rho, \hat{s}_t, \lambda_c))h_E(z_{jt})||_2 > \frac{e_1}{9}$ implies $\frac{1}{\sqrt{T}}||\sum_{j,t}(s_t - \hat{s}_t)||_2 > e_2$ thence letting $\mathbb{P}^* = \mathbb{P}(\cdot|n_t, x_t, \xi_t)$

$$
\begin{aligned}
&\mathbb{P}^*\Big(\frac{1}{\sqrt{T}}||\sum_{j,t}(D_j(\rho,s_t,\lambda_c)-D_j(\rho,\hat{s}_t,\lambda_c))h_E(z_{jt})||_2>\frac{e_1}{9}\Big)\leqslant \mathbb{P}^*\Big(\frac{1}{\sqrt{T}}||\sum_{j,t}(s_t-\hat{s}_t)||_2>e_2\Big)\\
&\leqslant \frac{J\sum_t \mathbb{E}^*(||s_t-\hat{s}_t||_2)}{e_2\sqrt{T}}=\frac{J\sum_t \mathbb{E}^*\left(\sqrt{\sum_j (s_{jt}-\hat{s}_{jt})^2}\right)}{e_2\sqrt{T}}\leqslant \frac{J\sum_t\sqrt{\sum_j \mathbb{E}^*\left((s_{jt}-\hat{s}_{jt})^2\right)}}{e_2\sqrt{T}}\\
&\leqslant \frac{J\sum_t\sqrt{\sum_j \mathbb{E}^*\left((\frac{1}{n_t}\sum_{i=1}^{n_t} y_{ijt}-\mathbb{E}^*(y_{ijt}))^2\right)}}{e_2\sqrt{T}}=\frac{J\sum_t\sqrt{\sum_j Var^*(\frac{1}{n_t}\sum_{i=1}^{n_t} y_{ijt})}}{e_2\sqrt{T}}\\
&\leqslant \frac{J\sum_t\sqrt{\sum_j \frac{1}{n_t}Var^*(y_{ijt})}}{e_2\sqrt{T}}\leqslant \frac{J^{3/2}}{e_2}\frac{1}{\sqrt{T}}\sum_t\frac{1}{\sqrt{n_t}}
\end{aligned}
$$

where Markov inequality, Jensen inequality, the fact that $y_{ijt}\in\{0;1\}$, that ε_{ijt} is iid extreme-value type 1 distributed across i, j and t, and the fact that n_t is iid and independent of all other variables have been used. Then taking the expectations and summing over N on both sides implies by Assumption E(i)

$$
\sum_{c=1}^{N}\mathbb{P}\Big(\frac{1}{\sqrt{T}}||\sum_{j,t}(D_j(\rho,s_t,\lambda_c)-D_j(\rho,\hat{s}_t,\lambda_c))h_E(z_{jt})||_2>\frac{e_1}{9}\Big)\leqslant\frac{J^{3/2}N}{e_2}\sqrt{T}\mathbb{E}(n_t^{-1/2})\underset{T\to\infty}{\to}0
$$

Next using continuity of D_j in $\tilde{\lambda}$ it must be that for any $e_1>0$ there exists some N such that $\forall\tilde{\lambda}\in\Lambda_{0,c}^N$ such that $||\tilde{\lambda}-\lambda_c||_2\leqslant r_N$ implies

$$
\frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho,s_t,\tilde{\lambda})-D_j(\rho,s_t,\lambda_c)|\;||h_E(z_{jt}||_2\leqslant\frac{e_1}{9}
$$

because $r_N\underset{N\to\infty}{\to}0$. By definition of the supremum it also implies that

$$
\underset{\tilde{\lambda}\in\Lambda_{0,c}}{sup}\;\frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho,s_t,\tilde{\lambda})-D_j(\rho,s_t,\lambda_c)|\;||h_E(z_{jt}||_2\leqslant\frac{e_1}{9}
$$

The contraposition is that

$$
\underset{\tilde{\lambda}\in\Lambda_{0,c}}{sup}\;\frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho,s_t,\tilde{\lambda})-D_j(\rho,s_t,\lambda_c)|\;||h_E(z_{jt}||_2>\frac{e_1}{9}
$$

implies $\forall\tilde{\lambda}\in\Lambda_{0,c}^N\quad||\tilde{\lambda}-\lambda_c||_2>r_N$ which is impossible by definition of $\Lambda_{0,c}^N$.

Consequently

$$\sum_{c=1}^{N}\mathbb{P}(\underset{\tilde{\lambda}\in\Lambda_{0,c}}{sup}\ \frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho,s_t,\tilde{\lambda})-D_j(\rho,s_t,\lambda_c)|\ ||h_E(z_{jt}||_2>\frac{e_1}{9})$$
$$\leqslant\ \sum_{c=1}^{N}\mathbb{P}(\cap_{\tilde{\lambda}\in\Lambda_{0,c}^N}||\tilde{\lambda}-\lambda_c||_2>r_N)=0$$

Similarly

$$\sum_{c=1}^{N}\mathbb{P}(\underset{\tilde{\lambda}\in\Lambda_{0,c}}{sup}\ \frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho,\hat{s}_t,\tilde{\lambda})-D_j(\rho,\hat{s}_t,\lambda_c)|\ ||h_E(z_{jt}||_2>\frac{e_1}{9})=0$$

(c) With the same arguments as in (b)

$$\mathbb{P}(\underset{\tilde{\lambda}}{sup}\ \frac{1}{\sqrt{T}}||\sum_{j,t}(D_j(\rho,\hat{s}_t,\tilde{\lambda})-D_j(\hat{\rho},\hat{s}_t,\tilde{\lambda}))h_E(z_{jt})||_2>\frac{e_1}{3})$$
$$\leqslant\ \sum_{c=1}^{N}\mathbb{P}(\frac{1}{\sqrt{T}}||\sum_{j,t}(D_j(\rho,\hat{s}_t,\lambda_c)-D_j(\hat{\rho},\hat{s}_t,\lambda_c))h_E(z_{jt})||_2>\frac{e_1}{9})$$
$$+\sum_{c=1}^{N}\mathbb{P}(\underset{\tilde{\lambda}\in\Lambda_{0,c}^N}{sup}\ \frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho,\hat{s}_t,\tilde{\lambda})-D_j(\rho,\hat{s}_t,\lambda_c)|\ ||h_E(z_{jt})||_2>\frac{e_1}{9})$$
$$+\sum_{c=1}^{N}\mathbb{P}(\underset{\tilde{\lambda}\in\Lambda_{0,c}^N}{sup}\ \frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\hat{\rho},\hat{s}_t,\lambda_c)-D_j(\hat{\rho},\hat{s}_t,\tilde{\lambda})|\ ||h_E(z_{jt})||_2>\frac{e_1}{9})$$
$$=\sum_{c=1}^{N}\mathbb{P}(\frac{1}{\sqrt{T}}||\sum_{j,t}(D_j(\rho,\hat{s}_t,\lambda_c)-D_j(\hat{\rho},\hat{s}_t,\lambda_c))h_E(z_{jt})||_2>\frac{e_1}{9})$$

where $D_j(\rho,s_t,\lambda_c)=C^{(H)}(\delta_0,s_t,x_{2t},f_0(\cdot|\lambda_c))$. D_j is $\mathcal{C}^\infty$ in $\rho\in(0;1)$, moreover $\rho_j(\delta_t,x_{2t},f_0(\cdot|\tilde{\lambda}))$ and $\hat{\rho}_j(\delta_t,x_{2t},f_0(\cdot|\tilde{\lambda}))$ are continuously differentiable in Λ_0. Therefore there exists some $e_2>0$ such that

$$\frac{1}{\sqrt{T}}\sum_{j,t}|D_j(\rho,\hat{s}_t,\lambda_c)-D_j(\hat{\rho},\hat{s}_t,\lambda_c)|\ ||h_E(z_{jt})||_2>\frac{e_1}{9}$$

implies $\underset{a\in\mathcal{B}}{sup}\ \frac{1}{\sqrt{T}}\sum_{j,t}||\rho(a,x_{2t},f_0(\cdot|\lambda_c)-\hat{\rho}(a,x_{2t},f_0(\cdot|\lambda_c))||_2>e_2$, and as $\mathcal{B}$ is compact we can cover it by $\tilde{N}$ closed balls $\mathcal{B}_b^{\tilde{N}}=\{a\in\mathcal{B}:||a-a_b||\leqslant r_{\tilde{N}}\}$

with $a_b \in \mathcal{B}$ for any $b = 1, \dots, \tilde{N}$ so that

$$\sum_{c=1}^{N} \mathbb{P}(\frac{1}{\sqrt{T}} \sum_{j,t} |D_j(\rho, \hat{s}_t, \lambda_c) - D_j(\hat{\rho}, \hat{s}_t, \lambda_c)| \; ||h_E(z_{jt})||_2 > \frac{e_1}{9})$$
$$\leqslant \sum_{c=1}^{N} \mathbb{P}(\sup_{a \in \mathcal{B}} \frac{1}{\sqrt{T}} \sum_{j,t} ||\rho(a, x_{2t}, f_0(\cdot|\lambda_c) - \hat{\rho}(a, x_{2t}, f_0(\cdot|\lambda_c))||_2 > e_2)$$
$$\leqslant \sum_{c,b} \mathbb{P}(\sup_{a \in \mathcal{B}_b^{\tilde{N}}} \frac{1}{\sqrt{T}} \sum_{j,t} ||\rho(a, x_{2t}, f_0(\cdot|\lambda_c) - \hat{\rho}(a, x_{2t}, f_0(\cdot|\lambda_c))||_2 > e_2)$$
$$= \sum_{c,b} \mathbb{P}(\frac{1}{\sqrt{T}} \sum_{j,t} ||\rho(a_b, x_{2t}, f_0(\cdot|\lambda_c) - \hat{\rho}(a_b, x_{2t}, f_0(\cdot|\lambda_c))||_2 > e_2)$$

where the last equality was obtained reusing arguments from (b). As a consequence let $F_{jt}(v) = \frac{exp(a_{bj} + x'_{2jt}v)}{1 + \sum_k exp(a_{bk} + x'_{2kt}v)}$ and $\mathbb{P}^*(\cdot) = \mathbb{P}(\cdot | x_t, \xi_t)$ then using Markov inequality and Cauchy-Schwarz inequality

$$\mathbb{P}^*(\frac{1}{\sqrt{T}} \sum_{j,t} ||\rho(a_b, x_{2t}, f_0(\cdot|\tilde{\lambda})) - \hat{\rho}(a_b, x_{2t}, f_0(\cdot|\tilde{\lambda}))||_2 > e_2)$$
$$\leqslant \frac{J \sum_t \mathbb{E}^*(||\hat{\rho}(a_b, x_{2t}, f_0(\cdot|\tilde{\lambda})) - \rho(a_b, x_{2t}, f_0(\cdot|\tilde{\lambda}))||_2)}{e_2 \sqrt{T}}$$
$$\leqslant \frac{J \sum_t \sqrt{\sum_j \mathbb{E}^* \left((\frac{1}{R} \sum_{r=1}^{R} F_{jt}(v_R) - \mathbb{E}^*(F_{jt}(v_R)))^2 \right)}}{e_2 \sqrt{T}} = \frac{J \sum_t \sqrt{\sum_j Var^*(\frac{1}{R} \sum_{r=1}^{R} F_{jt}(v_r))}}{e_2 \sqrt{T}}$$
$$\leqslant \frac{J^{3/2}}{e_2} \sqrt{\frac{T}{R}}$$

where the fact that v_r are iid draws from $f_0(\cdot|\tilde{\lambda})$ independent from all other variables has been used. It follows by taking the expectation and summing over N and $\tilde{N}$ that

$$\mathbb{P}(\sup_{\tilde{\lambda}} \frac{1}{\sqrt{T}} \sum_{j,t} |D_j(\rho, \hat{s}_t, \tilde{\lambda}) - D_j(\hat{\rho}, \hat{s}_t, \tilde{\lambda})| \; h_E(z_{jt})||_2 \underset{T \to \infty}{\to} 0$$

by Assumption E(i).

2. The 2nd statement is not formally proven as it largely builds on the proof of the 1st statement. To see why recall that

$$\hat{\Gamma}_T(\tilde{\theta}, h) - \Gamma_T(\tilde{\theta}, h) = \frac{1}{T} \sum_{jt} h(z_{jt}) \frac{\partial}{\partial \theta} (\widehat{\xi}(\tilde{\theta}) - \xi_{jt}(\tilde{\theta}))'$$

More precisely let $e'_j = (0 \; \dots \; 0 \quad \underbrace{1}_{\text{j-th coordinate}} \quad 0 \; \dots \; 0)$ then

$$\frac{\partial \xi_{jt}(\tilde{\theta})}{\partial \beta} = -x_{1jt}, \quad \frac{\partial}{\partial \lambda} \xi_{jt}(\tilde{\theta}) = -e'_j \left(\frac{\partial \rho(\delta_t(\tilde{\lambda}), x_{2t}, f_0(\cdot|\tilde{\lambda}))}{\partial \delta} \right)^{-1} \int \frac{exp(\delta_{jt}(\tilde{\lambda}) + x'_{2jt}v)}{1 + \sum_{k=1}^{J} exp(\delta_{kt}(\tilde{\lambda}) + x'_{2kt}v)} \frac{\partial}{\partial \lambda} f_0(v|\tilde{\lambda}) dv$$

Thus the columns of the matrix $\hat{\Gamma}_T(\tilde{\theta}, h) - \Gamma_T(\tilde{\theta}, h)$ associated to the derivative in β are equal to 0. Furthermore using an uniform continuity argument $\left|\frac{\partial\hat{\xi}_{jt}(\tilde{\theta})}{\partial\lambda} - \frac{\partial\xi_{jt}(\tilde{\theta})}{\partial\lambda}\right| > e_1$ is implied by $||\widehat{\delta}_t(\tilde{\lambda}) - \delta_t(\tilde{\lambda})||_2 > e_2$ for some $e_2 > 0$. Using the compactness of Λ_0 and Assumption E it is straightforward that $\underset{\tilde{\lambda}}{sup}\, ||\hat{\Gamma}_T(\tilde{\theta}, h) - \Gamma_T(\tilde{\theta}, h)||_2 \xrightarrow{P} 0$ for any h which satisfies the conditions in Assumption D.

3. The 3rd statement follows from the 1st. Indeed using Cauchy-Schwarz and properties of the supremum

$$
\begin{aligned}
\underset{\tilde{\theta}\in\Theta_0}{sup}\, |\hat{\mathcal{Q}}_T(\tilde{\theta}) - \mathcal{Q}_T(\tilde{\theta})| =& |(\hat{g}_T(\tilde{\theta}, h_E) - g_T(\tilde{\theta}, h_E))'\hat{W}(\hat{g}_T(\tilde{\theta}, h_E) - g_T(\tilde{\theta}, h_E)) \\
& - 2(\hat{g}_T(\tilde{\theta}, h_E) - g_T(\tilde{\theta}, h_E))'\hat{W} g_T(\tilde{\theta}, h_E))| \\
\leqslant & \underset{\tilde{\theta}\in\Theta_0}{sup}\, ||(\hat{g}_T(\tilde{\theta}, h_E) - g_T(\tilde{\theta}, h_E))||_2^2 \bar{\mu}(\hat{W}) \\
& + 2 \underset{\tilde{\theta}\in\Theta_0}{sup}\, ||(\hat{g}_T(\tilde{\theta}, h_E) - g_T(\tilde{\theta}, h_E))||_2 \underset{\tilde{\theta}\in\Theta_0}{sup}\, ||g_T(\tilde{\theta}, h_E))||_2 \bar{\mu}(\hat{W})
\end{aligned}
$$

where $\bar{\mu}(\cdot)$ maps a square matrix towards its maximum eigenvalue. By D(iv) and definition of the L_2 matrix norm, $\bar{\mu}(\hat{W}) \xrightarrow{P} \bar{\mu}(W)$. Then we apply Jennrich's ULLN: the data is iid, Θ_0 is compact, and $g_T(\tilde{\theta}, h_E) = \sum_j \xi_{jt}(f_0(\cdot|\tilde{\lambda}), \tilde{\beta}) h_E(z_{jt})$ has an enveloppe with finite absolute 1st moment because $\xi_{jt}(f_0(\cdot|\tilde{\lambda}), \tilde{\beta}) = \rho^{-1}(s_t, x_{2t}, \tilde{\lambda}) - x'_{1jt}\tilde{\beta}$ and $\rho^{-1}(\cdot)$ has a maximum because it is continuous and its input are in a compact and because $\tilde{\beta}$ is in a compact and x_{1jt} has finite 4th moments, see Assumption B; Thus by the CMT $\underset{\tilde{\theta}\in\Theta_0}{sup}\, ||g_T(\tilde{\theta}, h_E))||_2 \xrightarrow{P} \underset{\tilde{\theta}\in\Theta_0}{sup}\, ||g(\tilde{\theta}, h_E)||_2$; Finally using the 1st statement we have $||(\hat{g}_T(\tilde{\theta}, h_E) - g_T(\tilde{\theta}, h_E))||_2 \xrightarrow{P} 0$ therefore by the CMT

$$\underset{\tilde{\theta}\in\Theta_0}{sup}\, |\hat{\mathcal{Q}}_T(\tilde{\theta}) - \mathcal{Q}_T(\tilde{\theta})| \xrightarrow{P} 0$$

$\square$

B.4.5 Asymptotic Properties of the BLP estimator

Lemma 2.6 (Consistency of BLP estimator)
Given parametric assumption $\mathcal{F}_0$ and under assumptions B-E

$$\hat{\theta} \xrightarrow{P} \theta_0$$

Proof. We prove consistency using arguments for the consistency of M-estimators. For any $e_1 > 0$ such that $|\hat{\theta} - \theta_0| > e_1$ then by Assumption D(iii) there exists some $e_2 > 0$ such that $\mathcal{Q}(\hat{\theta}) - \mathcal{Q}(\theta_0) > e_2$ as θ_0 is the unique minimizer of the objective. Thence for any $e_1 > 0$, $\exists e_2 > 0$ such that

$$\begin{aligned}\mathbb{P}(|\hat{\theta} - \theta_0| > e_1) &\leqslant \mathbb{P}(\mathcal{Q}(\hat{\theta}) - \mathcal{Q}(\theta_0) > e_2)\\ &= \mathbb{P}(\hat{\mathcal{Q}}_T(\theta_0) - \mathcal{Q}(\theta_0) + \mathcal{Q}(\hat{\theta}) - \hat{\mathcal{Q}}_T(\hat{\theta}) + \hat{\mathcal{Q}}_T(\hat{\theta}) - \hat{\mathcal{Q}}_T(\theta_0) > e_2)\\ &\leqslant \mathbb{P}(\hat{\mathcal{Q}}_T(\theta_0) - \mathcal{Q}(\theta_0) + \mathcal{Q}(\hat{\theta}) - \hat{\mathcal{Q}}_T(\hat{\theta}) > e_2)\\ &\leqslant \mathbb{P}(\hat{\mathcal{Q}}_T(\theta_0) - \mathcal{Q}(\theta_0) > (1-\alpha)e_2) + \mathbb{P}(\mathcal{Q}(\hat{\theta}) - \hat{\mathcal{Q}}_T(\hat{\theta}) > \alpha e_2)\end{aligned}$$

where $\alpha \in (0;1)$, the 2nd inequality comes from the fact that $\hat{\mathcal{Q}}_T(\hat{\theta}) - \hat{\mathcal{Q}}_T(\theta_0)$ is almost surely negative by definition of $\hat{\theta}$, and the 3rd inequality is obtained by utilizing properties of indicator functions. Then by a direct implication of Lemma 2.5 the right-hand-side converges to 0.

$\square$

Lemma 2.7 (Asymptotic normality of BLP estimator)
Given parametric assumption $\mathcal{F}_0$, under assumptions B-E and under $H_0 : f \in \mathcal{F}_0$

$$\sqrt{T}(\hat{\theta} - \theta_0) = (\Gamma'(\theta_0, h_E)W\Gamma(\theta_0, h_E))^{-1}\sqrt{T}\Gamma'(\theta_0, h_E)Wg_T(\theta_0, h_E) + o_P(1)$$

Furthermore under $H_0; f \in \mathcal{F}_0$

$$\begin{aligned}\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} &\mathcal{N}(0, (\Gamma'(\theta_0, h_E)W\Gamma(\theta_0, h_E))^{-1}\Gamma'(\theta_0, h_E)W\Omega(\mathcal{F}_0, h_E)W\Gamma(\theta_0, h_E)\\ &(\Gamma'(\theta_0, h_E)W\Gamma(\theta_0, h_E))^{-1})\end{aligned}$$

Proof. We prove asymptotic normality using arguments from M-estimators asymptotics. From Taylor's Theorem there exists some $\tilde{\theta}$ such that $||\tilde{\theta} - \theta_0||_2 \leqslant ||\hat{\theta} - \theta_0||_2$ and

$$\begin{aligned}&\hat{g}_T(\hat{\theta}, h_E) = \hat{g}_T(\theta_0, h_E) + \hat{\Gamma}_T(\tilde{\theta}, h_E)(\hat{\theta} - \theta_0)\\ &\Rightarrow \sqrt{T}\hat{\Gamma}'_T(\hat{\theta}, h_E)\hat{W}\hat{g}_T(\hat{\theta}, h_E) = \sqrt{T}\hat{\Gamma}'_T(\hat{\theta}, h_E)\hat{W}\hat{g}_T(\theta_0, h_E) + \hat{\Gamma}'_T(\hat{\theta}, h_E)\hat{W}\hat{\Gamma}_T(\tilde{\theta}, h_E)\sqrt{T}(\hat{\theta} - \theta_0) = 0\\ &\Leftrightarrow \sqrt{T}(\hat{\theta} - \theta_0) = -\left(\hat{\Gamma}'_T(\hat{\theta}, h_E)\hat{W}\hat{\Gamma}_T(\tilde{\theta}, h_E)\right)^{-1}\sqrt{T}\hat{\Gamma}'_T(\hat{\theta}, h_E)\hat{W}\hat{g}_T(\theta_0, h_E)\end{aligned}$$

where the 1st implication is due to the FOC Assumption D(v). Then, we apply the CMT to $(A, B) \mapsto (A'BA)^{-1}A'B$ which is a continuous mapping if A and B are full rank so that when taking $A = \hat{\Gamma}_T(\hat{\theta}, h_E)$ and $B = \hat{W}$ we obtain:

$$\sqrt{T}(\hat{\theta} - \theta_0) = -(\Gamma'(\theta_0, h_E)W\Gamma(\theta_0, h_E))^{-1}\sqrt{T}\Gamma'(\theta_0, h_E)Wg_T(\theta_0, h_E) + o_P(1)$$

To prove that $plim\ \hat{\Gamma}_T(\hat{\theta}, h_E) = plim\ \hat{\Gamma}_T(\tilde{\theta}, h_E) = \Gamma(\theta_0, h_E)$ we make the following decomposition

$$\hat{\Gamma}_T(\hat{\theta}, h_E) - \Gamma(\theta_0, h_E) = \hat{\Gamma}_T(\hat{\theta}, h_E) - \Gamma_T(\hat{\theta}, h_E) + \Gamma_T(\hat{\theta}, h_E) - \Gamma(\hat{\theta}, h_E) + \Gamma(\hat{\theta}, h_E) - \Gamma(\theta_0, h_E)$$

where the 1st difference is $o_P(1)$ by Lemma 2.5, the 3rd difference is $o_P(1)$ by the CMT and the consistency of $\hat{\theta}$, see Lemma 2.6, and the 2nd difference is $o_P(1)$ by Jennrich's ULLN. The ULLN can be applied if and only if $\sum_j h_E(z_{jt}) \frac{\partial \xi_{jt}(\theta)}{\partial \theta}$ has an enveloppe with finite 1st absolute moments: $\xi_{jt}(\theta) = \rho^{-1}(s_t, x_{2t}, f_0(\cdot|\lambda)) - x'_{1jt}\beta$ and $\frac{\partial \xi_{jt}(\theta)}{\partial \beta} = x_{1jt}$ with x_{1jt} has finite moments of order 4 by Assumption B(iv), whereas $\frac{\partial \xi_{jt}(\theta)}{\partial \lambda} = \frac{\partial \rho^{-1}(s_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))}{\partial \lambda}$ and ρ^{-1} is C^∞ with arguments (s_t, x_{2t}, λ) which take values in a compact thus $\frac{\partial \rho^{-1}}{\partial \lambda}$ has bounds.

Thence $plim\ \hat{\Gamma}_T(\hat{\theta}, h_E) = plim\ \hat{\Gamma}_T(\tilde{\theta}, h_E) = \Gamma(\theta_0, h_E)$ which is full rank by Assumption D(ii), $plim\ \hat{W} = W$ which is full rank by Assumption D(iv), and by Lemma 2.5 $plim\ \sqrt{T}(\hat{g}_T(\theta_0, h_E) -\ g_T(\theta_0, h_E)) = 0$ so we can apply the aforementioned CMT and by the CLT which can be applied because $g(\theta_0, h_E) = 0$ under the null

$$\begin{aligned}\sqrt{T}(\hat{\theta} - \theta_0) &= -\left(\Gamma'(\theta_0, h_E) W \Gamma(\theta_0, h_E)\right)^{-1} \sqrt{T} \Gamma'(\theta_0, h_E) W g_T(\theta_0, h_E) + o_P(1) \\ &\xrightarrow{d} \mathcal{N}(0, (\Gamma'(\theta_0, h_E) W \Gamma(\theta_0, h_E))^{-1} \Gamma'(\theta_0, h_E) W \Omega(\mathcal{F}_0, h_E) W \Gamma(\theta_0, h_E) \\ &\quad (\Gamma'(\theta_0, h_E) W \Gamma(\theta_0, h_E))^{-1})\end{aligned}$$

□

B.4.6 Asymptotic distribution of the test statistic

Proof of Theorem 5.1

Proof. This proof leans heavily on the proof of Lemma 2.7. By Taylor's Theorem there exists $\tilde{\theta}$ such that $||\tilde{\theta} - \theta_0||_2 \leqslant ||\hat{\theta} - \theta_0||_2$

$$\begin{aligned}\sqrt{T}\hat{g}_T(\hat{\theta}, h_D) &= \sqrt{T}\hat{g}_T(\theta_0, h_D) + \hat{\Gamma}_T(\tilde{\theta}, h_D)\sqrt{T}(\hat{\theta} - \theta_0) \\ &= (I_{|h_D|0} \quad - \Gamma(\theta_0, h_D)(\Gamma'(\theta_0, h_D) W \Gamma(\theta_0, h_D))^{-1} \Gamma'(\theta_0, h_D) W) \sqrt{T} \begin{pmatrix} g_T(\theta_0, h_D) \\ g_T(\theta_0, h_E) \end{pmatrix} + o_P(1) \\ &\equiv (I_{|h_D|0} \quad G)\sqrt{T} \begin{pmatrix} g_T(\theta_0, h_D) \\ g_T(\theta_0, h_E) \end{pmatrix} + o_P(1)\end{aligned}$$

The second equality is obtained by relying on the proof of Lemma 2.7 to express $\sqrt{T}(\hat{\theta}-\theta_0)$ as a function of moments, by relying on Lemma 2.5 so that $plim\ \sqrt{T}\hat{g}_T(\theta_0,h_D) = plim\ \sqrt{T}g_T(\theta_0,h_D)$ and $plim\ \hat{\Gamma}_T(\tilde{\theta},h_D) = plim\ \Gamma_T(\theta_0,h_D)$, and by using the CMT.

- Under $H_0 : f \in \mathcal{F}_0$ then $\mathbb{E}\left[\sum_j h_D(z_{jt})\xi_{jt}(\theta_0)\right] = 0$ by LIE. So using the CLT and Slutsky's Lemma we obtain

$$\sqrt{T}\hat{g}_T(\hat{\theta},h_D) \xrightarrow{d} Z \sim \mathcal{N}(0,\Omega_0)$$

where

$$\Omega_0 = \begin{pmatrix} I_{|h_D|0} & G \end{pmatrix} \begin{pmatrix} \Omega(\mathcal{F}_0,h_D) & \Omega(\mathcal{F}_0,h_D,h_E) \\ \Omega(\mathcal{F}_0,h_D,h_E)' & \Omega(\mathcal{F}_0,h_E) \end{pmatrix} \begin{pmatrix} I_{|h_D|0} \\ G' \end{pmatrix}$$

with

$$\Omega(\mathcal{F}_0,h_D) = \mathbb{E}\Bigg[\Bigg(\sum_j \xi_{jt}(f(.|\lambda_0),\beta_0)h_D(z_{jt})\Bigg)\Bigg(\sum_j h_D(z_{jt})\xi_{jt}(f_0(.|\lambda_0),\beta_0)\Bigg)'\Bigg]$$

$$\Omega(\mathcal{F}_0,h_D,h_E) = \mathbb{E}\Bigg[\Bigg(\sum_j \xi_{jt}(f(.|\lambda_0),\beta_0)h_D(z_{jt})\Bigg)\Bigg(\sum_j h_E(z_{jt})\xi_{jt}(f_0(.|\lambda_0),\beta_0)\Bigg)'\Bigg]$$

$$G = -\Gamma(\theta_0,h_D)\left[\Gamma(\theta_0,h_E)'W\Gamma(\theta_0,h_E)\right]^{-1}\Gamma(\theta_0,h_E)'W$$

Thence by the continuous mapping theorem:

$$S(h_D,\mathcal{F}_0,\hat{\theta}) = \hat{g}_T(\hat{\theta},h_D)'\hat{\Sigma}\hat{g}_T(\hat{\theta},h_D) \xrightarrow{d} Z'\Sigma Z$$

$\square$

- Under $H_1' : \quad \mathbb{E}\left[\sum_j h_D(z_{jt})\xi_{jt}(f_0(\cdot|\lambda_0),\beta_0)\right] \neq 0$, we have by Lemma 2.5, by consistency of $\hat{\theta} \xrightarrow{P} \theta_0$ and the CMT:

$$\hat{g}_T(\hat{\theta},h_D) = g_T(\theta_0,h_D) + o_P(1)$$

Thus by Assumption D(iv) and the CMT

$$\frac{S(h_D,\mathcal{F}_0,\hat{\theta})}{T} \xrightarrow{P} \underbrace{\mathbb{E}\left[\sum_j h_D(z_{jt})\xi_{jt}(f_0(\cdot|\lambda_0),\beta_0)\right]'\Sigma\mathbb{E}\left[\sum_j h_D(z_{jt})\xi_{jt}(f_0(\cdot|\lambda_0),\beta_0)\right]}_{\kappa(h_D,\mathcal{F}_0,\theta_0)}$$

Under H_1', $\kappa(h_D, \mathcal{F}_0, \theta_0)$ is strictly positive because Σ is positive definite. Thence,

$$\begin{aligned}\forall q \in \mathbb{R} \quad \underset{T\to\infty}{lim} \; \mathbb{P}(S(h_D, \mathcal{F}_0, \hat{\theta}) > q) = \; & \underset{T\to\infty}{lim} \; \mathbb{P}\left(\frac{S(h_D, \mathcal{F}_0, \hat{\theta}) - q}{T} > 0\right) \\ &= \mathbb{P}(\kappa(h_D, \mathcal{F}_0, \theta_0) > 0) \\ &= 1\end{aligned}$$

where the 2nd equality holds because convergence in probability implies convergence in distribution.

$\square$

B.4.7 Application of Theorem 5.1 to the 2 polar cases

1. **Sargan-Hansen J test**
   If $h_D = h_E$, with W and Σ are set to be equal to the GMM 2-step optimal weighting matrix

$$\Sigma = W = \mathbb{E}\left[\left(\sum_j \xi_{jt}(f_0(\cdot|\lambda_0), \beta_0) h_E(z_{jt})\right)\left(\sum_j \xi_{jt}(f_0(\cdot|\lambda_0), \beta_0) h_E(z_{jt})\right)'\right]^{-1} = \Omega(\mathcal{F}_0, h_E)^{-1}$$

   Then under H_0:

$$S(h_D, \mathcal{F}_0, \hat{\theta}) \xrightarrow{d} \chi^2_{|h_E|_0 - |\theta|_0}$$

   Proof. By applying Theorem 5.1, we have:

$$S(h_D, \mathcal{F}_0, \hat{\theta}) \xrightarrow{d} Z'\Sigma Z \quad \text{with} \quad Z \sim \mathcal{N}(0, \Omega_0)$$

   If $h_D = h_E$ and $W = \Omega(\mathcal{F}_0, h_E)^{-1}$ then Ω_0 simplifies to

$$\begin{aligned}\Omega_0 &= \Omega(\mathcal{F}_0, h_E) - \Gamma(\theta_0, h_E)\left[\Gamma(\theta_0, h_E)'\Omega(\mathcal{F}_0, h_E)^{-1}\Gamma(\theta_0, h_E)\right]^{-1}\Gamma(\theta_0, h_E)' \\ &= \Omega(\mathcal{F}_0, h_E)^{1/2} M_{\Omega(\mathcal{F}_0, h_E)^{-1/2}\Gamma(\theta_0, h_E)} \Omega(\mathcal{F}_0, h_E)^{1/2}\end{aligned}$$

   with $M_{\Omega(\mathcal{F}_0, h_E)^{-1/2}\Gamma(\theta_0, h_E)} \equiv I_{|h_E|_0} - P_{\Omega(\mathcal{F}_0, h_E)^{-1/2}\Gamma(\theta_0, h_E)}$ is the orthogonal projection on the space orthogonal to $\Omega(\mathcal{F}_0, h_E)^{-1/2}\Gamma(\theta_0, h_E)$. Let $\tilde{Z} \sim \mathcal{N}(0, I_{|h_E|_0})$, we have by definition:

$$\begin{aligned}Z = \Omega(\mathcal{F}_0, h_E)^{1/2} M_{\Omega(\mathcal{F}_0, h_E)^{-1/2}\Gamma(\theta_0, h_E)} \tilde{Z} &\implies \Sigma^{1/2} Z = M_{\Omega(\mathcal{F}_0, h_E)^{-1/2}\Gamma(\theta_0, h_E)} \tilde{Z} \\ &\implies Z'\Sigma Z = \tilde{Z}' M_{\Omega(\mathcal{F}_0, h_E)^{-1/2}\Gamma(\theta_0, h_E)} \tilde{Z}\end{aligned}$$

The second line comes from symmetry and idempotence of $M_{\Omega(\mathcal{F}_0,h_E)^{-1/2}\Gamma(\theta_0,h_E)}$. Orthogonal projections have eigenvalues equal to either 0 or 1 with the number of eigenvalues equal to one corresponding to the rank of the space it projects into, which in our case is $|h_E| - |\theta|_0$. If we denote by V the matrix of eigenvectors of $M_{\Omega(\mathcal{F}_0,h_E)^{-1/2}\Gamma(\theta_0,h_E)}$ then note that $V'\tilde{Z} \sim \mathcal{N}(0, I_{|h_E|_0})$ so that

$$Z'\Sigma Z = \sum_{k=1}^{|h_E|_0-|\theta|_0} (V'\tilde{Z})_k^2 \sim \chi^2_{|h_E|_0-|\theta|_0}$$

□

2. **Non-overlapping h_D and h_E**
If Ω_0 is full rank and if the econometrician sets $\Sigma = \Omega_0^{-1}$, then our test statistic has the following asymptotic distribution under H_0:

$$S(h_T, \mathcal{F}_0, \hat{\theta}) \xrightarrow{d} \chi^2_{|h_D|_0}$$

One sufficient condition for Ω_0 being full rank is $(\xi_{jt}(f(\cdot|\lambda_0), \beta_0))_{j=1}^J$ is independent across j and $(h_E(z_{jt}), h_D(z_{jt}))$ not being perfectly colinear.

Proof. The asymptotic result is direct; $(\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0))_{j=1}^J$ being independent across j and $(h_E(z_{jt}), h_D(z_{jt}))$ not being perfectly colinear implies that

$$\Omega(\mathcal{F}_0, h_E, h_D) = \sum_j \mathbb{E}\left[\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0)^2 h_E(z_{jt}) h_D(z_{jt})'\right]$$

$$\Rightarrow \Omega_0 = \sum_j (I_{|h_D|_0} \quad G) Var\left(\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0)\begin{pmatrix} h_D(z_{jt}) \\ h_E(z_{jt}) \end{pmatrix}\right)\begin{pmatrix} I_{|h_D|_0} \\ G' \end{pmatrix}$$

Thus Ω_0 is positive definite because it is the sum of positive definite matrices. □

B.5 Properties of the MPI in the composite specification test: $f \in \mathcal{F}_0$

Proof of Proposition 3.3.

From Corollary **??**. Under Assumption A,

$$
\begin{aligned}
H_a: \ f \notin \mathcal{F}_0 \implies & \quad \mathbb{E}[\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0)|z_{jt}] \neq 0 \ \ a.s \\
\implies & \quad \mathbb{E}[\xi_{jt}(f_0(.|\lambda_0), \beta_0)|z_{jt}]^2 > 0 \ a.s \\
\implies & \quad \mathbb{E}\big[\mathbb{E}[\xi_{jt}(f_0(.|\lambda_0), \beta_0)|z_{jt}]^2\big] > 0 \\
\implies & \quad \mathbb{E}\big[\mathbb{E}[\xi_{jt}(f_0(.|\lambda_0), \beta_0)\mathbb{E}[\xi_{jt}(f_0(.|\lambda_0)|z_{jt}]|z_{jt}]\big] > 0 \\
\implies & \quad \mathbb{E}\big[\xi_{jt}(f_0(.|\lambda_0), \beta_0)\mathbb{E}[\xi_{jt}(f_0(.|\lambda_0)|z_{jt}]\big] > 0 \\
\implies & \ \forall \alpha \neq 0 \ \ H_1': \ \mathbb{E}\big[\xi_{jt}(f_0(.|\lambda_0), \beta_0) \underbrace{\alpha\mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]}_{h_D^*(z_{jt})}\big] \neq 0
\end{aligned}
$$

From Theorem 5.1, under assumptions B-E,

$$
H_a \implies \forall q \in \mathbb{R}^+, \ \ \mathbb{P}(S(h_D^*, \mathcal{F}_0, \hat{\theta}) > q) \to 1
$$

□

C Additional results and comments

C.1 Literature on the identification of the distribution of RC

In this section, we briefly summarize some recent findings on the identification of random coefficients in multinomial choice models has been extensively studied in the literature. In their seminal paper, Berry and Haile (2014) shows the identification of the demand functions ρ in a framework that encompasses the BLP model but their result does not entail identification of the random coefficients' distribution per se. To achieve their identification result, they require a completeness condition on the instruments as well as additional conditions (eg: connected substitutes) to ensure invertibility of the demand functions. They also need to impose that at least one of the product characteristic has a coefficient that is not random and that is equal to 1. Notice that in BLP model, the structure implied by the logit shock guarantees invertibility of the demand functions.

Fox et al. (2012) provides conditions under which the distribution of random coefficients is identified in a mixed logit model with micro-level data and no endogeneity. Their identification result requires continuous characteristics in x_{2t} and rules out interaction terms (eg polynomial terms of x_{2jt}). Moreover, their result is restricted to distributions of random coefficients with a compact support - excluding for instance a normally distributed random coefficient.

Fox and Gandhi (2011) investigates the identification of the joint distribution of random coefficients v_i and idiosyncratic shocks ε_{ijt} in aggregate demand models without endogeneity. They also consider a setting where endogeneity is introduced in a very restrictive way. They show identification under a special regressor assumption and finite support of the unobserved heterogeneity. The special regressor assumption assumes that a variable in x_{1t} has full support and has an associated coefficient that is either 1 or -1. This special regressor assumption is very common in the literature on the identification of random coefficients (see Ichimura and Thompson (1998), Berry and Haile (2009), Matzkin (2007) and Lewbel (2000)). Their framework does not nest the standard BLP model as ϵ_{ijt} and v_i are both assumed to have a finite support but it is more general in other dimensions. They do not exploit the logit distributional assumption on ε_{ijt}, they do not impose independence between v_i and ε_{ijt}, their identification argument can be extended to the case where multiple goods are purchased.

In a setting much closer to ours, Dunker et al. (2022) studies the identification of the distribution of random coefficients in endogenous aggregate demand models which includes the BLP model as a special case (in particular, no parametric assumption is made on the idiosyncratic shock ε_{ijt}). They make a clever use of the Radon transform to identify f. The price they have to incur for flexibility is that they need to make stringent assumptions on the product characteristics: variables in x_t are required to be continuous and to satisfy a joint full support assumption. The idea is to exploit the variation in the covariates in order to trace out the distribution of rc f. Unfortunately, these requirements are rarely met in real data sets.

In contrast to the rest of the literature, Wang (2022) adopts all the parametric assumptions in the standard BLP model and looks for the set of minimal assumptions under which the distribution of random coefficients is identified. This approach allows him to obtain sufficient conditions which are much less stringent than the rest of the literature (no special regressor assumption, no full support assumption, no continuity assumption). To be more specific, he shows that if the demand functions are identified on an open set of $\mathbb{R}^J$[38], then the distribution of random coefficients is identified. His proof astutely exploits the real analytic property of the demand functions[39].

[38] which can be achieved using Theorem 1 in Berry and Haile (2014)

[39] In particular, the real analytic property yields that the local identification of ρ on $\mathcal{D} \subset \mathbb{R}^J$ implies identification of ρ on $\mathbb{R}^J$. From global identification of ρ, he is then able to show that the random coefficients' distribution is identified under a simple rank condition on x_{2t}

C.2 Feasible MPI: conditional expectation

C.3 Choice of the large-T asymptotics

In this paper, we study the asymptotics of our test when the number of markets T grows to infinity. We could also develop an asymptotic theory with J growing to infinity and T staying fixed but there are many arguments against it. First, from an economic stand point, it's hard to conceptually think of a market with a very large number of products and some form of independence across products. Second, from a theoretical point of view there is a tension between the identification of demand which require all market shares to be strictly positive, see Berry and Haile (2014), and the large market asymptotics which require all market shares to tend to 0 as J grows to infinity, see Berry et al. (2004). At the same time it is well established that a many (weak) instruments problem can easily occur in a BLP model with a fixed number of markets and many products especially when using the traditional BLP instruments, see Armstrong (2016).

Consequently, only markets with perfect competition and a careful choice of instruments could somehow fit the assumptions necessary for the BLP model to yield consistent estimators and valid tests with large J. Yet in the majority of empirical IO papers the markets have imperfect competition, sometimes oligopolies, and use the traditional BLP instruments. Thus we establish our theory with a large number of independent markets, which is a natural setting for empirical IO papers and which is not plagued with the aforementioned theoretical problems.

C.4 Construction of the interval instruments in practice

We now provide more details on how to construct the interval instruments in practice. The procedure to construct the interval instruments is as follows:

1. Given $(\mathcal{F}_0, \hat{W}, h_E)$, the researcher derives the BLP estimator $\hat{\theta}$
2. Then the researcher chooses L points $(v_l)_{l=1}^L \in \mathbb{R}^L$ in the presumed support of $f_0(\cdot|\hat{\lambda})$.
3. Finally, the researcher can construct a set of L interval instruments based on the approximations of the MPI that we develop in sections 4.2 and 4.1.

   - Global approximation: $\{\pi_{j,l}(z_{jt})\}_{l=1,\dots,L}$ interval instruments which are such that:

$$\mathbb{E}\left[\Delta_j(s_t, x_{2t}, f_0, f_a)|z_{jt}\right] \approx \log\left(\sum_{l=1}^{L} \omega_l \; \pi_{j,l}(z_{jt})\right) \text{ with } \pi_{j,l}(z_{jt}) = \frac{\frac{\exp(x_{2jt}' v_l)}{1+\sum_{k=1}^J \exp\left\{\hat{\delta}_{kt}^0 + x_{2kt}' v_l\right\}}}{\int_{\mathbb{R}^{K_2}} \frac{\exp(x_{2jt}' v)}{1+\sum_{k=1}^J \exp\left\{\hat{\delta}_{jt}^0 + x_{2kt}' v\right\}} f_0(v) dv}$$

with $\hat{\delta}_t^0$ the linear projection of δ_t^0 on z_{jt} (or a carefully chosen subset of z_{jt}).

- Local approximation: $\{\bar{\pi}_{j,l}(z_{jt})\}_{l=1,\dots,L}$ interval instruments such that

$$\mathbb{E}[\Delta_j(s_t, x_{2t}, f_0, f_a)|z_{jt}] \approx \sum_{l=1}^{L} \bar{\omega}_l \ \bar{\pi}_{j,l}(z_{jt})$$

$$\text{with } \bar{\pi}_{j,l}(z_{jt}) = \left(\frac{\partial \rho(\hat{\delta}_t^0, x_{2t}, f_0)}{\partial \delta}\right)^{-1} \left[\frac{\exp(\hat{\delta}_t^0 + x_{2t} v_l)}{1 + \sum_{k=1}^{J} \exp\left\{\hat{\delta}_{kt}^0 + x'_{2kt} v_l\right\}} - \rho_j(\hat{\delta}_t^0, x_{2t}, f_0)\right]$$

with $\hat{\delta}_t^0$ the linear projection of δ_t^0 on z_{jt} (or a carefully chosen subset of z_{jt}).

Choice of the L points in the domain of f_a The researcher doesn't know a priori the support of the true density f_a. Thus, he/she must choose points in the domain of definition of f_a. If this choice coincides with points of the support where $|f_0(\cdot|\lambda_0) - f_a|$ is large, then this choice generates more informative instruments. In practice, one can take points in the high density regions of $f_0(\cdot|\lambda_0)$ (eg if $\mathcal{F}_0$ is the Gaussian family, then one can take points around the mean λ_0). The choice of of the number of instruments N obeys a usual bias variance tradeoff. On the one hand, a large L allows to better approximate the MPI and thus increases the detection ability of the instruments. On the other hand, it is well-known that a larger number of instruments can induce finite sample bias and can distort asymptotic distributions of estimators and tests such as the over-identification test[40]; For these reasons we advise not to use too few or too many interval instruments, in our simulations and application we use between 10 and 20 instruments. We leave a formal analysis of the optimal choice of L and of the general approximations properties of the interval instruments for future work.

C.5 Feasible MPIs for estimation

As for the global approximation we derived in section 4.2, it is straightforward to show that for any candidate $f_0(\cdot|\lambda_0)$, we can rewrite this approximation of the non-linear part of the MPI as follows:

$$\mathbb{E}[\Delta_j(s_t, x_{2t}, f_0(\cdot|\lambda_0), f_a)|z_{jt}] \approx \log\left(\sum_{l=1}^{L} \bar{\bar{\omega}}_l(f_0(\cdot|\lambda_0), f_a) \ \hat{\bar{\bar{\pi}}}_{j,l}(z_{jt})\right) \text{ with } \hat{\bar{\bar{\pi}}}_{j,l}(z_{jt}) = \frac{\exp(x'_{2jt} v_l)}{1 + \sum_{k=1}^{J} \exp\left\{\hat{\bar{\bar{\delta}}}_{jt}^a + x_{2jk} v_l\right\}}$$

$$\textit{and } \bar{\bar{\omega}}_l(f_0, f_a) = \frac{\bar{\omega}_l(f_a)}{\int_{\mathbb{R}^{K_2}} \frac{\exp(x'_{2jt} v)}{1 + \sum_{k=1}^{J} \exp\left\{\delta_{jt}^0 + x'_{2jk} v\right\}} f_0(\cdot|\lambda_0)(v) dv}$$

[40] see Roodman (2009) for a review on the effect of many possibly weak moments on estimation and testing

with $\hat{\hat{\delta}}^a_{jt}$ projected first stage estimates of δ^a_{jt}, which can be obtained, for example, under the logit specification. $\hat{\hat{\pi}}_{j,l}(z_{jt})$ don't depend on f_0 and can be used for estimation.

C.6 Estimation procedure when the distribution of RC is a mixture

In this section, we present a procedure to estimate the BLP model when the distribution of RC is parametrized as a mixture. Namely, we perform the estimation under $H_0 : f \in \mathcal{F}_0$ with $\mathcal{F}_0$ the family of Gaussian mixtures with L components. The pdf of a Gaussian mixture writes as follows:

$$\forall x \in \mathbb{R}\ ,\ f_0(x|\lambda_0) = \sum_{l=1}^{L} p_{l0} f_l(x|\lambda_{l0}) \qquad \sum_{l=1}^{L} p_{l0} = 1 \qquad L \geqslant 1$$

where $f_{l0}(\cdot|\lambda_{l0})$ is the pdf of a $\mathcal{N}(\mu_{l0}, \sigma^2_{l0})$.

As long as the means are different ($\mu_{l0} \neq \mu_{l'0}\ \forall l \neq l'$), the gaussian mixture is uniquely characterized by the vector $\lambda_0 = (p_{10}, \ldots, p_{L0}, \mu_{10}, \ldots, \mu_{L0}, \sigma^2_{10}, \ldots, \sigma^2_{L0})$ up to permutations of indexes[41]. The objective of our procedure is to estimate the parameters of the model $\theta_0 = (\beta_0, \lambda_0)$ where λ_0 characterizes the mixture. In general, the problem of estimating a density by a mixture is solved through the use of the well-known Expectation-Maximization (EM) algorithm. In our case, the application of this algorithm is made difficult by two main obstacles. First, we do not observe directly the random coefficients. Second, we do not have individual choice data which would have enabled us to construct a likelihood as in Train (2008). As an alternative, we propose to adapt the BLP estimation procedure to estimate the parameters of a mixture of gaussians instead of the single normal distribution. The mixture affects the derivation of the market shares. The random coefficient v_i is now a gaussian mixture. Hence, $v_i = \sum_{l=1}^{L} \mathbf{1}\{D_i = l\} v_{il}$ where $(v_{il})_{i=1}^n$ are iid and have density $f_{l0}(\cdot|\lambda_{l0})$ known up to λ_{l0} for $l = 1, \ldots, L$, and where $(D_i)_{i=1}^n$ are iid categorically distributed with pmf $\mathbb{P}(D_i = l) = p_{l0}$. For all market t and product j, the demand functions are as follows:

$$\begin{aligned} \rho_j(\delta_t, x_{2t}, f_0(.|\lambda_0)) &= \mathbb{P}(j \text{ chosen in market } t \text{ by } i|x_{1t}, x_{2t}, \xi_t) \\ &= \int_{\mathbb{R}} \frac{\exp\{x'_{1jt}\beta_0 + x'_{2jt}v + \xi_{jt}\}}{1 + \sum_{j'=1}^{J} \exp\{x'_{1j't}\beta_0 + x'_{2j't}v + \xi_{j't}\}} f_0(v|\lambda_0)dv \\ &= \sum_{l=1}^{L} p_{l0} \int_{\mathbb{R}} \frac{\exp\{\delta_{jt} + x'_{2jt}v\}}{1 + \sum_{j'=1}^{J} \exp\{\delta_{j't} + x'_{2j't}v\}} f_{l0}(v|\lambda_{l0})dv \end{aligned}$$

[41] If for some $l \neq l'$ we have $\mu_{l0} = \mu_{l'0}$ then the Gaussian mixture becomes observationally equivalent to an infinite number of other Gaussian mixtures

Reparametrization The parameter λ associated with the mixture consists of the means, the standard deviation and the probability of each component. As highlighted by Ketz (2019) in the simple Gaussian case, the way we parametrize the model can greatly affect the asymptotic properties of the estimator as well as the quality of the estimation. In particular, he shows that the standard deviations σ should be reparametrized in order to avoid boundaries issues when σ close to 0. We follow this parametrization and perform the minimization with respect to $\{(+/-)\sqrt{\sigma_l}\}_{l=1}^{L}$ instead and $(\sigma_l)_{l=1}^{L}$ directly. An additional difficulty in the case of mixtures concerns the estimation of the probabilities associated to each component. These probabilities must all be between 0 and 1 and their sum must be equal to 1. To smoothly integrate these constraints, we perform the optimization with respect to $\gamma = (\gamma_2, \dots, \gamma_L)$ with $p = (p_1, p_2, \dots, p_L) = (\frac{1}{1+\sum_{l=2}^{L} exp(\gamma_l)}, \frac{exp(\gamma_2)}{1+\sum_{l=2}^{L} exp(\gamma_l)}, \dots, \frac{exp(\gamma_L)}{1+\sum_{l=2}^{L} exp(\gamma_l)})$.

Estimation details Apart from the modification in the computation of the market shares and the new parametrization of the model, the estimation procedure with a mixture follows closely the traditional one and the parameters of interest are estimated by minimizing a GMM criterion. Let $\mathcal{Q}(\theta)$ the GMM objective function:

$$\mathcal{Q}(\theta) = \hat{\xi}(\theta)' h_E(Z) W h_E(Z)' \hat{\xi}(\theta)$$

We now describe the derivation of the Gradient that we provide to the minimization program.

$$\frac{\partial \mathcal{Q}}{\partial \theta} = 2 \left[\frac{\partial \hat{\xi}(\theta)}{\partial \theta} \right]' h_E(Z) W h_E(Z)' \hat{\xi}(\theta)$$

Where $\frac{\partial \hat{\xi}(\theta)}{\partial \beta} = -x_1$ and where by the implicit function theorem we have $\hat{\rho}_j(\delta_t, x_{2t}, \lambda) - s_{jt} = 0 \ \ \forall j, t$ which implies:

$$\frac{\partial \hat{\xi}(\theta)}{\partial \lambda} = \frac{\partial \hat{\delta}(\theta)}{\partial \lambda} = - \left[\frac{\partial \hat{\rho}(\delta, x_2, \lambda)}{\partial \delta} \right]^{-1} \frac{\partial \hat{\rho}(\delta, x_2, \lambda)}{\partial \lambda}$$

- $\frac{\partial \rho}{\partial \delta}$ is a $JT \times JT$ diagonal by block matrix such that:

$$\frac{\partial \rho_j\left(\delta_t, x_{2t}, \lambda\right)}{\partial \delta_{kt}} = \begin{cases} \sum_l p_l \int \mathcal{T}_{jlt}(v) \left(1 - \mathcal{T}_{klt}(v)\right) \phi_l(v) dv & if \quad j = k \\ - \sum_l p_l \int \mathcal{T}_{jlt}(v) \mathcal{T}_{klt}(v) \phi_l(v) dv & if \quad j \neq k \end{cases}$$

with $\mathcal{T}_{jlt}(v) \equiv \frac{exp\{\delta_{jt} + x'_{2jt} v_l\}}{1+\sum_{j'=1}^{J} exp\{\delta_{j't} + x'_{2j't} v_l\}}$

- $\frac{\partial \rho}{\partial \lambda}$ is a $JT \times (3L-1)$ matrix such that:

$$\frac{\partial \rho_j\left(\delta_t, x_{2t}, \lambda\right)}{\partial \mu_l} = p_l \int \mathcal{T}_{jlt}\left(x_{2jt} - \sum_{j'} \mathcal{T}_{j'lt} x_{2j't}\right) \phi(v) dv$$
$$\frac{\partial \rho_j\left(\delta_t, x_{2t}, \lambda\right)}{\partial \sigma_l} = p_l \int \mathcal{T}_{jlt}\left(x_{2jt} - \sum_{j'} \mathcal{T}_{j'lt} x_{2j't}\right) v\phi(v) dv$$
$$\frac{\partial \rho_j\left(\delta_t, x_{2t}, \lambda\right)}{\partial \gamma_l} = \sum_{l'=1}^{L} \zeta(l, l') \int \mathcal{T}_{jlt}$$

With $\zeta(l,l') = \frac{-exp(\gamma_l)}{1+\sum_{k\neq 1} exp(\gamma_k)} \times \frac{exp(\gamma_{l'})}{1+\sum_{k\neq 1} exp(\gamma_k)} + \mathbf{1}\{l = l'\} \frac{exp(\gamma_l)}{1+\sum_{k\neq 1} exp(\gamma_k)} = -p_l \times p_{l'} + \mathbf{1}\{l = l'\} p_l$

C.7 Properties of the feasible approximations of the MPI

So far we have studied the properties of the MPI, which is an ideal instrument that cannot be derived in practice. Nevertheless, in light of the previous results, the MPI provides a useful upper bound on the power that can be reached using our specification test. More precisely, the asymptotic slope reached by the MPI can be interpreted as a power envelope on our specification test. Ideally, we want our specification test, with the approximated MPIs as instruments, to achieve slopes close to the ones reached by the MPI. For the sake of exposition, let us assume homoskedasticity. We now distinguish 2 situations.

First, we consider the case where the econometrician tests H_0 against the true alternative $\bar{H}_a : (f, \beta) = (f_a, \beta_a)$. This situation is not interesting in practice as the econometrician usually doesn't know the true alternative and doesn't want to specify an alternative. Nevertheless, it illustrates that in this specific case, we can (in theory) derive a consistent estimator of the MPI. Indeed, in this particular case, we can directly derive an analytical expression for the correction term $\Delta_{0,a}^{\xi_{jt}}$ either using its definition or the expression in 4.2. Next, we must to compute the conditional expectation of our the correction term with respect to z_{jt}. This step is quite challenging because the dimension of z_{jt} is large and because the correction term is heavily non-linear and non-separable with respect to the endogenous variables. In theory, a solution is to perform a Sieve non-parametric estimation of the conditional mean and under standard regularity conditions recover a consistent estimator of $\mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]$. Unfortunately, the rate of converge will be extremely slow given the dimension of z_{jt} and we don't recommend to do this in practice. Instead, we suggest to use the global approximation and to "exogenize" the endogenous variables by projecting them on the space spanned by a relevant subset of z_{jt}. As we show in the Appendix, this strategy yields an estimator which converges faster to a first order approximation of the MPI.

Second, we consider the more realistic situation where the econometrician tests H_0 against an unspecified alternative. This is the situation of interest in this paper. In this case, we use the interval instruments that we developed in section 4 as an approximation of the MPI. Due to the different layers of approximations which intervene in the construction of these instruments and the absence of knowledge of f_a, it is quite difficult to establish conditions under which these instruments can reach the optimal slope of the MPI. A thorough analysis of the properties of these instruments is beyond the scope of this paper and may constitute an interesting starting point for future research. In the Appendix, we present a preliminary investigation on the theoretical properties of the local approximation. In spite of the lack of theoretical analysis, our Monte Carlo exercises show that the interval instruments perform really well in finite sample.

Approximation properties of the interval instruments The interval instruments, ie the approximation of the MPI denoted as $\hat{h}_T^*$, work well in practice in the sense that they yield a valid test which is powerful. However it is difficult to prove that the speed of divergence of our test when using them is as large as the speed of divergence when using the true MPI without further assumptions. As described in the previous subsection there are 3 levels of approximation to the MPI: First, only the 1st order of the expansion of the difference between the true error and generated error is considered, another term $\mathcal{R}_0$ remains[42]; Second, the conditional expectation with respect to the full set of instruments is approximated via projections; Third, the integral which appears within this 1st order approximation is estimated via a Riemann sum of N points. Consequently, if $\mathcal{R}_0$ is negligible, if N is very large, and if projecting the difference between the generated error and the true error is equivalent to taking its conditional expectation with respect to the instruments, then the slope $C_{\hat{h}_T^*}$ is equal to $C_{h_T^*}$. This result is summarized in the following proposition:

Proposition 3.1

Under Assumption B and C, and assuming strict homoskedasticity $\mathbb{E}(\xi_{t0}\xi_{t0}'|z_t) = I_J$ then under $H_1 : f \notin \mathcal{F}_0$ there exists some sequence $(\alpha)_{l=1}^N$ such that

$$\hat{h}_T^*(z_{jt})'\alpha + err_{jt} + e\tilde{r}r_{jt} \underset{N\to+\infty}{\longrightarrow} h_T^*(z_{jt})$$

almost surely, for some errors $err_{jt} \in \mathbb{R}^N$ and $e\tilde{r}r_{jt} \in \mathbb{R}$ such that $e\tilde{r}r_{jt} \underset{N\to+\infty}{\overset{as}{\longrightarrow}} 0$. As a

[42]we have obtained the formula of the approximation of the difference between the true error and the generated error up to the second order

consequence

$$C_{\hat{h}^*_T} = \mathbb{E}\left(\sum_j \alpha' \hat{h}^*_T(z_{jt}) \hat{h}^*_T(z_{jt})' \alpha\right) + error + e\tilde{r}ror$$

for some $error \in \mathbb{R}$ *and some* $e\tilde{r}ror \in \mathbb{R}$ *such that* $e\tilde{r}ror \xrightarrow[N\to+\infty]{as} 0$. *If* $error = 0$ *then*

$$C_{\hat{h}^*_T} \underset{N\to+\infty}{\longrightarrow} C_{h^*_T} = \mathbb{E}\left(\sum_j \mathbb{E}(\tilde{\Delta}_{jt}(s_t, x_{2t}, \mathcal{F}_0, f)|z_{jt})^2\right)$$

To further comment on this result err_{jt} ($error$) corresponds to the first and second errors of approximations of the MPI described above and $e\tilde{r}r_{jt}$ ($e\tilde{r}ror$) corresponds to the third; On the other hand $(\alpha)_{l=1}^N$ is a sequence of integration weights whose empirical mean converge to 0. In addition there are 2 conditions necessary for $error = 0$. The first and most important one is that $\mathcal{R}_0$ shoud be close to 0, in other words the 1st order approximation should explain most of the difference between the generated error and the true error. The second condition for $error$ to be close to 0 is very likely to be satisfied in practice: We need to be able to approximate well the conditional expectation with respect to z_{jt} of the 1st order approximation of $\tilde{\Delta}$. As noted by Reynaert and Verboven (2014), because most product characteristics are uncorrelated with the unobserved product characteristics, using a Sieve estimator of the conditional expectation or a more practical method as is described in our paper or theirs does not seem to make a lot of difference. If these two conditions are satisfied then err_{jt} is small and therefore $error$ is small.

C.7.1 Proof of Proposition 3.1

Using the strict homoskedasticity assumption then from ?? we know that

$$C_{h^*_T} = \mathbb{E}(\mathbb{E}(\tilde{\Delta}_t(\mathcal{F}_0, f)|z_t)' \mathbb{E}(\tilde{\Delta}_t(\mathcal{F}_0, f)|z_t)) = \mathbb{E}(h^*_T(z_t)' h^*_T(z_t))$$

Our goal is therefore to prove that under some conditions

$$\underset{N\to+\infty}{lim} C_{\hat{h}^*_T} = C_{h^*_T}$$

and we do so in four steps: First, we prove that there exists some (err_1, err_2, err_3) such that

$$\hat{h}^*_T(z_t)'\alpha + err_{1t} + err_{2t} + err_{3t} \underset{N\to+\infty}{\to} h^*_T(z_t)$$

Second, we show that $err_{3t} \underset{N\to+\infty}{\to} 0$ almost surely; Third, we show that there exists some $\tilde{h}^*_T(z_t)$ and some $(error, e\tilde{r}ror)$ such that

$$C_{\hat{h}^*_T} = C_{\tilde{h}^*_T}, \qquad C_{h^*_T} = \alpha' \mathbb{E}(\tilde{h}^*_T(z_t)' \tilde{h}^*_T(z_t))\alpha + error + e\tilde{r}ror$$

and $e\tilde{r}ror \xrightarrow[N\to+\infty]{as} 0$; Fourth we conclude. We prove each point in order:

- Denote and recall

$$\eta_{jt} = \int \frac{exp(\hat{\delta}^0_{jt} + x'_{2jt}v)}{1+\sum_{k=1}^J exp(\hat{\delta}_0 + x'_{2kt}v)}(f(v) - f_0(v|\lambda_0))dv$$

$$\hat{\eta}_{jt,l} = \frac{exp(\hat{\delta}^0_{jt} + x'_{2jt}v_l)}{1+\sum_{k=1}^J exp(\hat{\delta}_0 + x'_{2kt}v_l)}$$

$$\Rightarrow \hat{\eta}'_{jt}\alpha = \sum_l \alpha_l \frac{exp(\hat{\delta}^0_{jt} + x'_{2jt}v_l)}{1+\sum_{k=1}^J exp(\hat{\delta}_0 + x'_{2kt}v_l)}$$

$$M_t(\cdot) = x_{1t}\left(\mathbb{E}\left[\sum_j x_{1jt}h_E(z_{jt})'\right]W\mathbb{E}\left[\sum_j h_E(z_{jt})x'_{1jt}\right]\right)^{-1}\mathbb{E}\left[\sum_j x_{1jt}h_E(z_{jt})'\right]W\mathbb{E}\left[\sum_j h_E(z_{jt})\right.$$

$$\hat{M} = \hat{x}_1\left[x'_1 h_E(z)(h_E(z)'h_E(z))^{-1}h_E(z)'x_1\right]^{-1}x'_1 h_E(z)\hat{W}h_E(z)'$$

$$M^{-1}_{t,\partial\rho} = \left(\frac{\partial\rho(\delta^0_t, x_{2t}, f_0(\cdot|\lambda_0))}{\partial\delta}\right)^{-1}$$

$$\hat{M}^{-1}_{t,\partial\rho} = \left(\frac{\partial\rho(\hat{\delta}^0_t, \hat{x}_{2t}, f_0(\cdot|\hat{\lambda}))}{\partial\delta}\right)^{-1}$$

where $(\hat{x}_1, \hat{x}_2, \hat{\delta}^0)$ are transformations of (x_1, x_2, δ^0) (for instance their projection on the instruments) as described in **??**. Then define $\hat{M}_{\partial\rho}$ of dimension $(J\times T)\times(J\times T)$ which is block diagonal with T blocks of dimension $J\times J$ equal to $\hat{M}^{-1}_{t,\partial\rho}$, and define $\hat{\eta}$ which is the stacked versions of $\hat{\eta}_{jt}$. Consequently

$$\begin{aligned}
h^*_T(z_t) = \mathbb{E}(\tilde{\Delta}(\mathcal{F}_0, f)|z_t) &= \mathbb{E}\left((id - M_t)\Delta(s_t, x_{2t}, \mathcal{F}_0, f)|z_t\right) \\
&= \mathbb{E}\left((id - M_t)(M^{-1}_{t,\partial\rho}\eta_t + \mathcal{R}_0)|z_t\right) \\
\hat{h}^*_T(z_t)\alpha &= A_t\hat{h}_T(z)\alpha \\
&= A_t(I_{J\times T} - \hat{M})\hat{\Delta}'_N\alpha \\
&= A_t(I_{J\times T} - \hat{M})\hat{M}^{-1}_{\partial\rho}\hat{\eta}\alpha
\end{aligned}$$

where A_t is the matrix which picks the J observations in t, ie A_t is a $J\times(J\times T)$ matrix of zeros except the block from column $(J-1)t+1$ to Jt which is equal to I_J. In other words

$$\begin{aligned}
h^*_T(z_t) &= \hat{h}^*_T(z_t)\alpha + \mathbb{E}\left((id - M_t)\mathcal{R}_0|z_t\right) \\
&\quad + \left[\mathbb{E}((id - M_t)M^{-1}_{t,\partial\rho}\eta_t|z_t) - A_t(I_{J\times T} - \hat{M})\hat{M}^{-1}_{\partial\rho}\lim_{N\to+\infty}\hat{\eta}\alpha\right] \\
&\quad + \left[\lim_{N\to+\infty}(\hat{h}^*_T(z_t)\alpha) - \hat{h}^*_T(z_t)\alpha\right] \\
&\equiv \hat{h}^*_T(z_t)\alpha + err_{1t} + err_{2t} + err_{3t}
\end{aligned}$$

- Next clearly if $(v_l, c_{l,N})_{l=1}^N$ are chosen so that $\forall l \;\; v_l$ is in the support of $f(\cdot) - f_0(\cdot|\hat{\lambda})$ and $v_{l+1} - v_{l+1} = c_{l,N} \underset{N\to+\infty}{\to} 0$ then the Riemann sum

$$\hat{\eta}'_{jt}\alpha = \sum_l \alpha_l \frac{exp(\hat{\delta}^0_{jt} + x'_{2jt}v_l)}{1 + \sum_{k=1}^J exp(\hat{\delta}_0 + x'_{2kt}v_l)} = \sum_l \frac{c_{l,N}}{N}(f(v_l) - f_0(v_l|\hat{\lambda}))\frac{exp(\hat{\delta}^0_{jt} + x'_{2jt}v_l)}{1 + \sum_{k=1}^J exp(\hat{\delta}_0 + x'_{2kt}v_l)}$$

converges to $\int \frac{exp(\hat{\delta}^0_{jt} + x'_{2jt}v)}{1 + \sum_{k=1}^J exp(\hat{\delta}_0 + x'_{2kt}v)}(f(v) - f_0(v|\hat{\lambda}))dv$ almost surely when $N \to +\infty$, see arguments for the convergence of Riemann sums. This integral exists by Assumption C(ii) and implicitly by Assumption **??**(i)-(ii). Therefore for any t $err_{3t} \underset{N\to+\infty}{\to} 0$ almost surely, which corresponds to $\tilde{err}_t$ in the proposition.

- Next for a fixed N, by Assumption **??** using the LLN and the CMT, there exists some $\tilde{h}^*_T(z_t)$ which is the "probability limit" of $\hat{h}^*_T(z_t)$ in the sense that

$$\frac{1}{T}S(\hat{h}^*_T, \mathcal{F}_0, \theta_0) = \frac{1}{T}S(\tilde{h}^*_T, \mathcal{F}_0, \theta_0) + o_P(1), \qquad \tilde{h}^*_T(z_t) = (id - \tilde{M})\tilde{M}^{-1}_{t,\partial\rho}\tilde{\eta}_t, \qquad C_{\hat{h}^*_T} = C_{\tilde{h}^*_T}$$

where $BLP(\cdot|z_t)$ is the best linear projection operator and

$$\tilde{M}_t(\cdot) = \tilde{x}_{1t}\left(\mathbb{E}\left[\sum_j x_{1jt}h_E(z_{jt})'\right]W\mathbb{E}\left[\sum_j h_E(z_{jt})x'_{1jt}\right]\right)^{-1}\mathbb{E}\left[\sum_j x_{1jt}h_E(z_{jt})'\right]W\mathbb{E}\left[\sum_j h_E(z_{jt})\cdot\right]$$

$$\tilde{M}_{t,\partial\rho} = \left(\frac{\partial\rho(\tilde{\delta}^0_t, \tilde{x}_{2t}, \mathcal{F}_0, \lambda_0)}{\partial\delta}\right)^{-1}$$

$$\tilde{\eta}_{jt,l} = \frac{exp(\tilde{\delta}^0_{jt} + \tilde{x}'_{2jt}v_l)}{1 + \sum_k exp(\tilde{\delta}^0_{kt} + \tilde{x}'_{2kt}v_l)}$$

$$\tilde{\delta}^0_t = \delta^0_t - BLP(\delta^0_t|z_t)$$

$$\tilde{x}_{1t} = x_{1t} - BLP(x_{1t}|z_t)$$

$$\tilde{x}_{2t} = x_{2t} - BLP(x_{2t}|z_t)$$

As a consequence $h^*_T(z_t)$ rewrites

$$\begin{aligned} h^*_T(z_t) = {} & \tilde{h}^*_T(z_t)\alpha + \mathbb{E}\left((id - M_t)\mathcal{R}_0|z_t\right) \\ & + \left[\mathbb{E}((id - M_t)M^{-1}_{t,\partial\rho}\eta_t|z_t) - (id - \tilde{M}_t)(\tilde{M}^{-1}_{t,\partial\rho}\lim_{N\to+\infty}\tilde{\eta}_t\alpha)\right] \\ & + \left[\lim_{N\to+\infty}(\tilde{h}^*_T(z_t)\alpha) - \tilde{h}^*_T(z_t)\alpha\right] \\ \equiv {} & \tilde{h}^*_T(z_t)\alpha + \tilde{err}_{1t} + \tilde{err}_{2t} + \tilde{err}_{3t} \\ \Rightarrow C_{h^*_T(z_t)} \equiv {} & \alpha'\mathbb{E}(\tilde{h}^*_T(z_t)'\tilde{h}^*_T(z_t))\alpha + \tilde{error} + error \end{aligned}$$

where $e\tilde{r}\tilde{o}r$ is a function of $e\tilde{r}r_{3t}$ and therefore converges to 0 almost surely as $N \to +\infty$ and $error$ is a function of $\tilde{h}^*_T(z_t)$, $e\tilde{r}r_{1t}$ and $e\tilde{r}r_{2t}$.

- From the previous point if $e\tilde{r}r_{1t} = e\tilde{r}r_{2t} = 0$, ie

$$\mathcal{R}_0 = 0, \qquad \left[\mathbb{E}((id - M_t)M^{-1}_{t,\partial\rho}\eta_t|z_t) - (id - \tilde{M}_t)(\tilde{M}^{-1}_{t,\partial\rho} \underset{N\to+\infty}{lim} \tilde{\eta}_t\alpha)\right] = 0$$

Then $h^*_T(z_t) \underset{N\to+\infty}{=} \tilde{h}^*_T(z_t)\alpha$ thus $C_{h^*_T} \underset{N\to+\infty}{=} C_{\tilde{h}^*_T\alpha} = C_{\hat{h}^*_T\alpha}$. Finally using the properties of best linear projections it can be shown that $C_{\hat{h}^*_T} = C_{\tilde{h}^*_T} \geqslant C_{\tilde{h}^*_T\alpha} = C_{\hat{h}^*_T\alpha}$ so that $\underset{N\to+\infty}{lim} C_{\hat{h}^*_T} = C_{h^*_T\alpha}$ because $C_{h^*_T\alpha}$ also constitutes an upper bound on $C_{\hat{h}^*_T}$. Indeed

$$\begin{aligned}
C_{\tilde{h}^*_T} &= \mathbb{E}\left(\tilde{\Delta}_t(\mathcal{F}_0,f)'\tilde{h}^*_T(z_t)\right)\mathbb{E}\left(\tilde{h}^*_T(z_t)'\xi_{0t}\xi'_{0t}\tilde{h}^*_T(z_t)\right)^{-1}\mathbb{E}\left(\tilde{h}^*_T(z_t)'\tilde{\Delta}_t(\mathcal{F}_0,f)\right)\\
&= \mathbb{E}\left(\tilde{\Delta}_t(\mathcal{F}_0,f)'\tilde{h}^*_T(z_t)\right)\mathbb{E}\left(\tilde{h}^*_T(z_t)'\tilde{h}^*_T(z_t)\right)^{-1}\mathbb{E}\left(\tilde{h}^*_T(z_t)'\tilde{\Delta}_t(\mathcal{F}_0,f)\right)\\
&\geqslant C_{\tilde{h}^*_T\alpha} = \mathbb{E}\left(\tilde{\Delta}_t(\mathcal{F}_0,f)'\tilde{h}^*_T(z_t)\right)\alpha\mathbb{E}\left(\alpha'\tilde{h}^*_T(z_t)'\tilde{h}^*_T(z_t)\alpha\right)^{-1}\alpha\mathbb{E}\left(\tilde{h}^*_T(z_t)'\tilde{\Delta}_t(\mathcal{F}_0,f)\right)
\end{aligned}$$

where the first second equality is due to the fact that we assume strict exogeneity $\mathbb{E}(\xi_{0t}\xi_{0t}|z_t) = I_J$, and the inequality is due to the fact that the best linear projection of $\tilde{\Delta}_t(\mathcal{F}_0,f)$ on the subspace $\tilde{h}^*_T(z_t)\alpha$ always has lower second moment compared to the best linear projection of $\tilde{\Delta}_t(\mathcal{F}_0,f)$ on the space $\tilde{h}^*_T(z_t)$.

D Monte Carlo experiments

D.1 Counterfactuals under an alternative distribution

Expressions for price and cross-price elasticities as a function of p_1 in the simulation exercise presented in section 6.2

- Price elasticity:

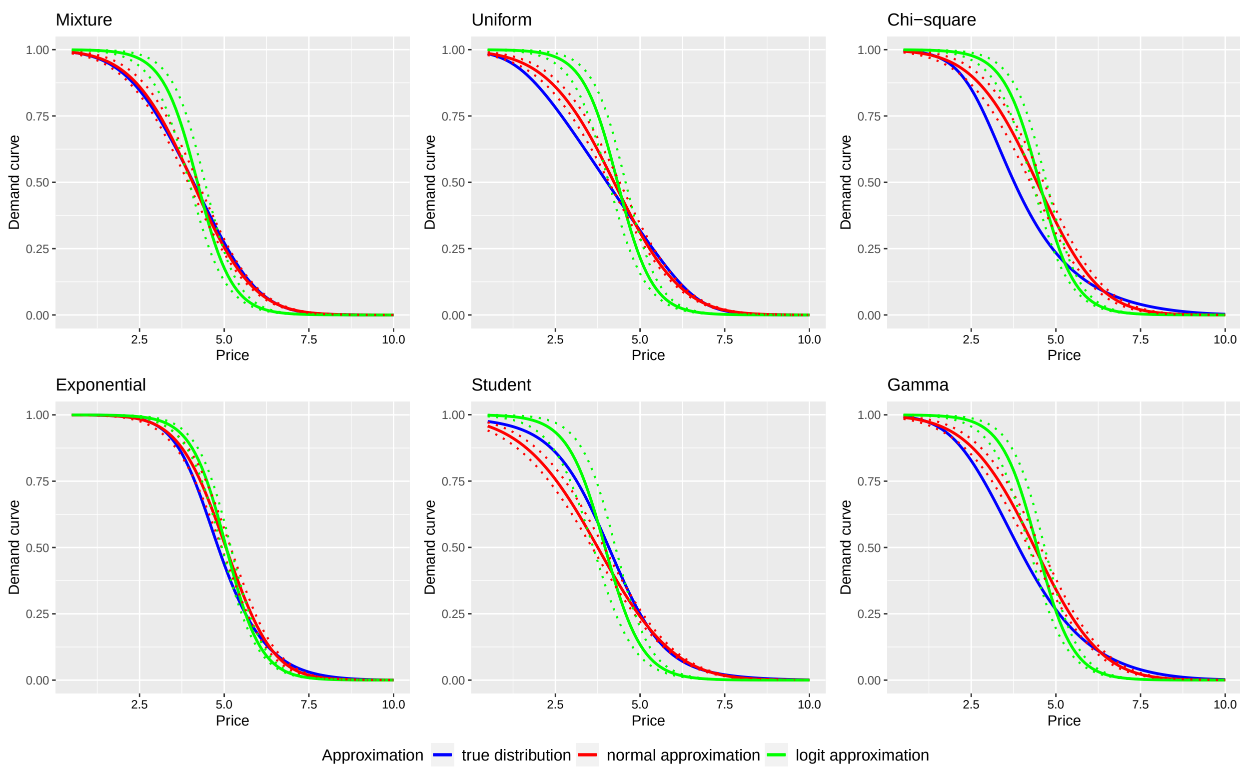
$$\xi_1 = \frac{p_1}{s_1}\frac{\partial s_1}{\partial p_1} = \int -\alpha\left(1 - \frac{\exp\{u_{i1}\}}{1+\sum_{j'\in\{1,2\}}\exp\{u_{ij'}\}}\right)\frac{exp\{u_{i1}\}}{1+\sum_{j'\in\{1,2\}}\exp\{u_{ij'}\}}f_\theta(v)dv\phi(\alpha)d\alpha$$

- Cross price elasticity:

$$\xi_{2/1} = \frac{p_1}{s_2}\frac{\partial s_2}{\partial p_1} = \int \alpha\left(\frac{\exp\{u_{i1}\}}{1+\sum^J_{j'\in\{1,2\}}\exp\{u_{ij'}\}}\right)\frac{\exp\{u_{i2}\}}{1+\sum^J_{j'\in\{1,2\}}\exp\{u_{ij'}\}}f_\theta(v)dv\phi(\alpha)d\alpha$$

Here, we plot the demand functions generates by the different approximations of the true densities

Figure 9: Demand function



D.2 Finite sample performance of the test

Practical implementation of the test For each setting, we estimate the model for 1000 replications. Minimization is performed with nloptr (algorithm: NLOPT-LD-LBFGS). We provide an analytical gradient. The Threshold for the outer loop is 1e-9 while the threshold for the inner loop is 1e-13. We use squarem and a C++ implementation for the computation of the market shares to speed up the contraction. We also parallelize the contraction over markets using 7 independent cores. Now we formally describe the instruments included in each test.

Instruments

- J(1): differentiation instruments + exogenous characteristics (polynomial terms) + cost shifters (15 instruments/ degrees of over-identification:8)

- I(1): first stage instruments: instruments J(1). testing instruments: Interval Instruments: 7 instruments. Points chosen as follows: $\{\hat{\mu},(\hat{\mu}+k(max(0.25,\hat{\sigma})),k(max(0.25,\hat{\sigma}))\}$ $(for\ k=1,2,3)$

- J(2): first stage: instruments: instruments J(1). second stage instruments: optimal instruments (approximation of $\mathbb{E}\left[\frac{\partial\rho_j^{-1}(s_t,x_{2t},\lambda)}{\partial\lambda}\middle|z_t\right]$)+ exogenous characteristics (polynomial terms) + cost shifters (12 instruments)

- I(2): first stage instruments: instruments J(2). Testing instruments: Interval Instruments: 7 instruments. Points chosen as follows: $\{\hat{\mu},(\hat{\mu}+k(max(0.25,\hat{\sigma})),k(max(0.25,\hat{\sigma}))\}$ $(for\ k=1,2,3)$

Power against local alternatives We now assess the local power properties of our test by assuming that the random coefficient v_i is distributed according to a local alternative. Namely, we assume $v_i\sim\left(1-\frac{1}{\sqrt{T}}\right)\mathcal{N}(2,1)+\frac{1}{\sqrt{T}}Y$ where Y is an alternative distribution including exponential, Chi-square, Student, Uniform. We ensure that Y has mean 2 and variance 1. The results are reported in 13. First, we can observe that except for the uniform local alternative, our test appears to have non-trivial power against all the other local alternatives. For the exponential and chi-square distributions, it is clear that our test with interval instruments outperforms the Sargan-J test with traditional instruments. For the student local alternative, the results seem quite unstable for small sample sizes but as T increases, interval instruments also seem to perform better. For the uniform alternative, it appears that we don't have power against this local alternative.

Table 13: Empirical power, local alternatives (1000 replications)

Number of markets	T=50				T=100				T=200			
Test type	J	I	J	I Local	J	I Local	J	I	J	I Local	J	I Local
Exponential	0.266	0.704	0.227	0.677	0.222	0.869	0.272	0.868	0.236	0.982	0.394	0.975
Chi-square	0.217	0.219	0.134	0.174	0.13	0.167	0.096	0.151	0.099	0.171	0.086	0.15
Student	0.212	0.139	0.33	0.436	0.115	0.115	0.127	0.093	0.082	0.13	0.134	0.312
Uniform	0.198	0.1	0.126	0.074	0.107	0.062	0.095	0.051	0.073	0.049	0.084	0.044

D.3 Finite sample performance of Interval instruments for estimation

Practical implementation of the estimation procedure To assess the performance of our instruments in estimating the non-linear parameters with a flexible distribution of random coefficients, we simulate data with a distribution of random coefficients following a mixture of gaussians and we estimate the parameters of this mixture. For each setting, we estimate the model for 1050 replications. We select the replications with an objective function below a certain threshold (in order to avoid local minima). Minimization is performed with nloptr (algorithm: NLOPT-LD-LBFGS). We provide an analytical gradient, which we describe subsequently. The Threshold for the outer loop is 1e-9 while the threshold for the inner loop is 1e-13. We use squarem and a C++ implementation for the computation of the market shares to speed up the contraction. We also parallelize the contraction over markets using 7 independent core. Before we formally define the different sets of instruments, let us present the estimation procedure when the distribution of random coefficients is assumed to be a mixture.

Instruments Now we formally describe the instruments present in each different sets used for estimation

- Differentiation instruments: differentiation instruments + exogenous characteristics (polynomial terms) + cost shifters (20 instruments)

- Optimal instruments are computed in two stages. The first stage instruments consist of differentiation instruments and exogenous characteristics (polynomial terms). Second stage instruments consist of polynomial terms of exogenous characteristics and the approximation of optimal instruments proposed in Reynaert and Verboven (2014) (approximation of $\mathbb{E}\left[\frac{\partial \rho_j^{-1}(s_t, x_{2t}, \lambda)}{\partial \lambda}\middle| z_t\right]$). The set called optimal instruments includes 15 instruments.

- Interval Instruments are computed in two stages. The first stage instruments consist of differentiation instruments and exogenous characteristics (polynomial terms). Second stage instruments are the interval instruments couples with some exogenous characteristics. A total of 23 instruments. The points in the support to compute the interval instruments are chose as follows: we take equally spaced points in the interval $\{\beta_{3L} - 0.5(\beta_{3H} - \beta_{3L}), \beta_{3H} + 0.5(\beta_{3H} - \beta_{3L})\}$.

Comparison of the performance between the different sets of instruments We now report the mean biases and the empirical $\sqrt{MSE}$ of the estimates for each set of instruments and for different sample sizes. We also plot the distributions of estimates for the non-linear parameters for the different sets of instruments. First, we plot the distribution of estimates obtained when the set of differentiation instruments from Gandhi and Houde (2019) is used with a sample of $T = 200$ markets and $J = 12$ products. We observe that despite a relatively large sample, the differentiation instruments perform rather poorly in estimating the non-linear parameters associated with the mixture of Gaussians. In particular, the estimates of the standard deviation parameters associated to each component are very dispersed and a large portion of the estimates are bunched at zero. Second, we plot the distribution of non-linear estimates obtained with the optimal instruments from Reynaert and Verboven (2014). They tend to perform better than the differentiation instruments as we can see that the estimates are more concentrated around the true value. Yet, it is important to emphasize that the optimal instruments display large failure rates caused by perfect colinearity of the instruments. We report the percentage of replications that subject to perfect colinearity issues for each sample size $(39\%, 34\%, 31\%, 26\%, 23\%)$. Finally, we plot the distribution of estimates for the non linear parameters when we use the interval instruments developed in section **??**. It appears clearly that the interval instruments yield a more concentrated distribution of estimates than the two other sets of instruments. For the sake of conciseness, we do not report the results with a mixture with 3 components but the observations we make with two components are even more exacerbated.

Table 14: Estimation mixture with “differentiation” instruments (1000 replications)

	Parameter	β_0	α	β_1	β_2	β_{3L}	σ_{3L}	β_{3H}	σ_{3H}	p_L
Sample size	true	2	-2	1.5	1	-2	0.5	4	0.5	0.25
T=50, J=12	bias	-0.12	0.022	-0.016	-0.018	0.214	0.184	-0.022	-0.045	0.027
	$\sqrt{MSE}$	0.308	0.06	0.215	0.215	0.633	0.734	0.281	0.35	0.075
T=50, J=20	bias	-0.064	0.011	-0.01	-0.011	0.189	0.347	0.022	-0.081	0.025
	$\sqrt{MSE}$	0.231	0.044	0.165	0.166	0.566	0.887	0.184	0.291	0.059
T=100, J=12	bias	-0.058	0.01	-0.012	-0.012	0.233	0.226	0.02	-0.066	0.027
	$\sqrt{MSE}$	0.204	0.041	0.147	0.148	0.592	0.703	0.256	0.305	0.072
T=100, J=20	bias	-0.04	0.006	-0.007	-0.007	0.198	0.423	0.047	-0.101	0.025
	$\sqrt{MSE}$	0.165	0.032	0.117	0.116	0.552	0.89	0.164	0.27	0.055
T=200, J=12	bias	-0.038	0.007	-0.003	-0.003	0.184	0.167	0.011	-0.049	0.019
	$\sqrt{MSE}$	0.152	0.03	0.11	0.11	0.466	0.601	0.176	0.262	0.053

Table 15: Estimation mixture with "Optimal" instruments(1000 replications)

	Parameter	β_0	α	β_1	β_2	β_{3L}	σ_{3L}	β_{3H}	σ_{3H}	p_L
Sample size	true	2	-2	1.5	1	-2	0.5	4	0.5	0.25
T=50, J=12	bias	-0.09	0.016	-0.012	-0.013	0.076	0.059	0.026	-0.111	0.01
	$\sqrt{MSE}$	0.296	0.057	0.234	0.232	0.361	0.483	0.212	0.281	0.036
T=50, J=20	bias	-0.046	0.007	0	0.001	0.074	0.11	0.028	-0.089	0.01
	$\sqrt{MSE}$	0.225	0.044	0.178	0.176	0.328	0.563	0.163	0.228	0.033
T=100, J=12	bias	-0.041	0.007	-0.004	-0.003	0.054	0.037	0.019	-0.066	0.007
	$\sqrt{MSE}$	0.202	0.039	0.157	0.158	0.279	0.4	0.154	0.211	0.028
T=100, J=20	bias	-0.029	0.004	-0.003	-0.003	0.074	0.107	0.033	-0.074	0.01
	$\sqrt{MSE}$	0.153	0.03	0.126	0.124	0.311	0.52	0.129	0.194	0.034
T=200, J=12	bias	-0.029	0.005	-0.001	-0.001	0.026	0.011	0.021	-0.061	0.004
	$\sqrt{MSE}$	0.136	0.026	0.111	0.111	0.184	0.313	0.113	0.172	0.018

Table 16: Estimation mixture with Global Interval instruments(1000 replications)

	Parameter	β_0	α	β_1	β_2	β_{3L}	σ_{3L}	β_{3H}	σ_{3H}	p_L
Sample size	true	2	-2	1.5	1	-2	0.5	4	0.5	0.25
T=50, J=12	bias	-0.154	0.029	-0.043	-0.045	0.017	0	-0.045	0.004	0.005
	$\sqrt{MSE}$	0.341	0.067	0.257	0.258	0.277	0.391	0.227	0.259	0.024
T=50, J=20	bias	-0.092	0.017	-0.02	-0.021	0.013	0.042	-0.018	-0.003	0.004
	$\sqrt{MSE}$	0.245	0.048	0.19	0.19	0.248	0.415	0.166	0.22	0.021
T=100, J=12	bias	-0.07	0.013	-0.017	-0.019	0.004	-0.012	-0.027	0.005	0.002
	$\sqrt{MSE}$	0.2	0.039	0.161	0.161	0.167	0.282	0.157	0.201	0.013
T=100, J=20	bias	-0.047	0.008	-0.006	-0.007	-0.009	-0.005	-0.008	-0.009	0.001
	$\sqrt{MSE}$	0.158	0.031	0.13	0.129	0.115	0.264	0.115	0.169	0.005
T=200, J=12	bias	-0.039	0.007	-0.004	-0.003	-0.006	-0.027	-0.015	-0.001	0.001
	$\sqrt{MSE}$	0.141	0.027	0.109	0.109	0.088	0.219	0.108	0.164	0.003

Table 17: Estimation mixture with Local Interval instruments(1000 replications)

	Parameter	β_0	α	β_1	β_2	β_{3L}	σ_{3L}	β_{3H}	σ_{3H}	p_L
Sample size	true	2	-2	1.5	1	-2	0.5	4	0.5	0.25
T=50, J=12	bias	-0.134	0.025	-0.023	-0.024	-0.006	-0.005	-0.039	-0.001	0.003
	$\sqrt{MSE}$	0.307	0.059	0.26	0.259	0.251	0.34	0.214	0.244	0.019
T=50, J=12	bias	-0.084	0.016	-0.024	-0.025	0.019	0.033	-0.023	0.01	0.003
	$\sqrt{MSE}$	0.245	0.047	0.188	0.186	0.228	0.38	0.15	0.184	0.018
T=50, J=12	bias	-0.075	0.015	-0.018	-0.016	0	0	-0.028	0.007	0.001
	$\sqrt{MSE}$	0.199	0.039	0.159	0.16	0.127	0.225	0.143	0.164	0.005
T=50, J=12	bias	-0.039	0.007	-0.011	-0.011	-0.003	0.004	-0.01	0.004	0.001
	$\sqrt{MSE}$	0.162	0.032	0.129	0.129	0.104	0.226	0.103	0.125	0.004
T=50, J=12	bias	-0.037	0.007	-0.008	-0.007	0.002	-0.007	-0.016	0.006	0.001
	$\sqrt{MSE}$	0.136	0.026	0.11	0.109	0.091	0.174	0.099	0.123	0.003

Figure 10: Distribution of estimates for non-linear parameters with "Differentiation" instruments ($T = 200$, $J = 12$)

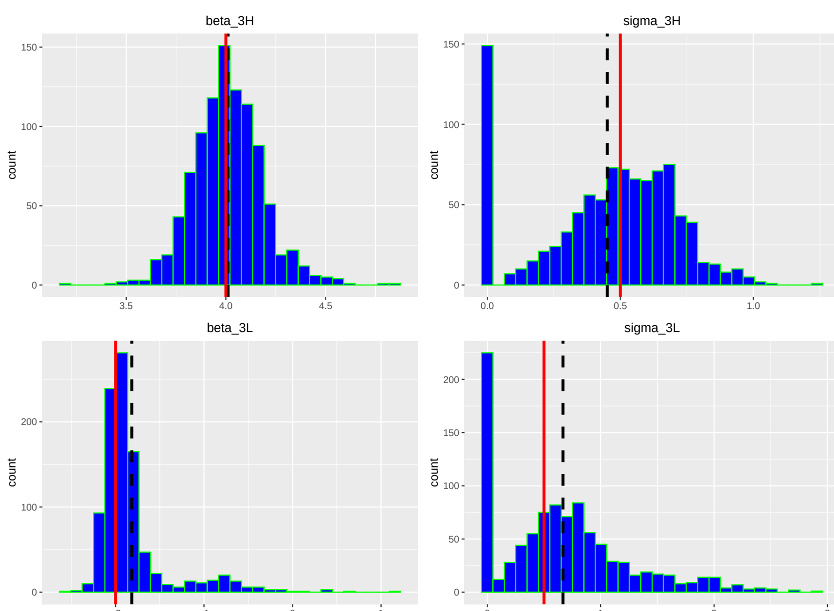


Figure 11: Distribution of estimates for non-linear parameters with "Optimal" instruments ($T = 200, J = 12$)

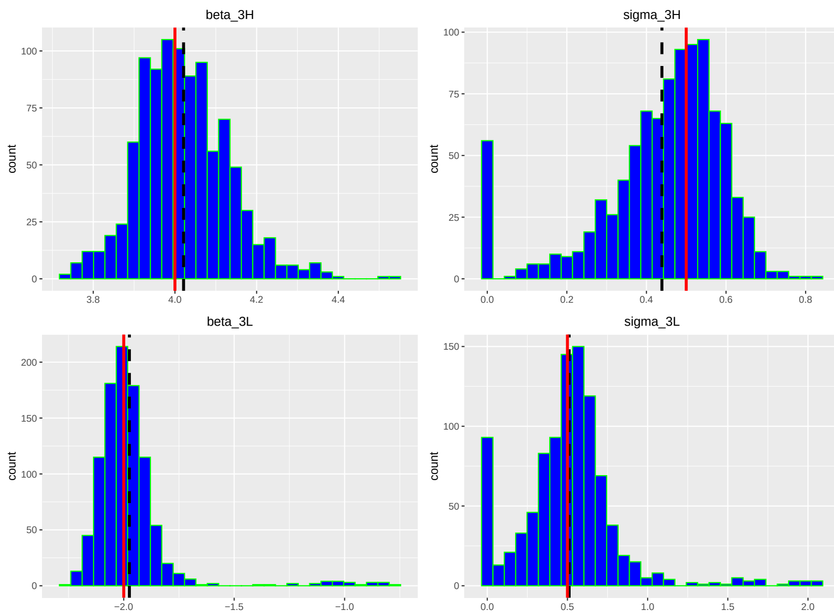


Figure 12: Distribution of estimates for non-linear parameters with "Global Interval" instruments ($T = 200, J = 12$)

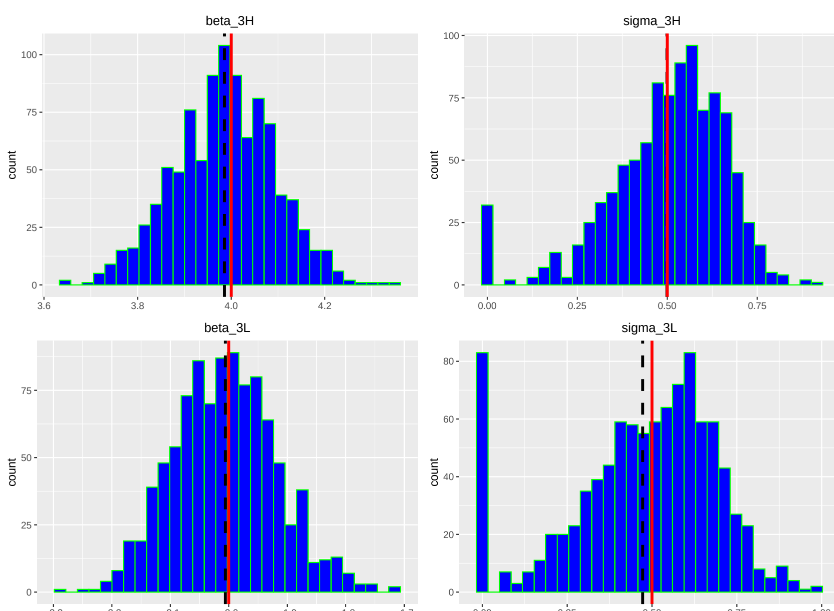
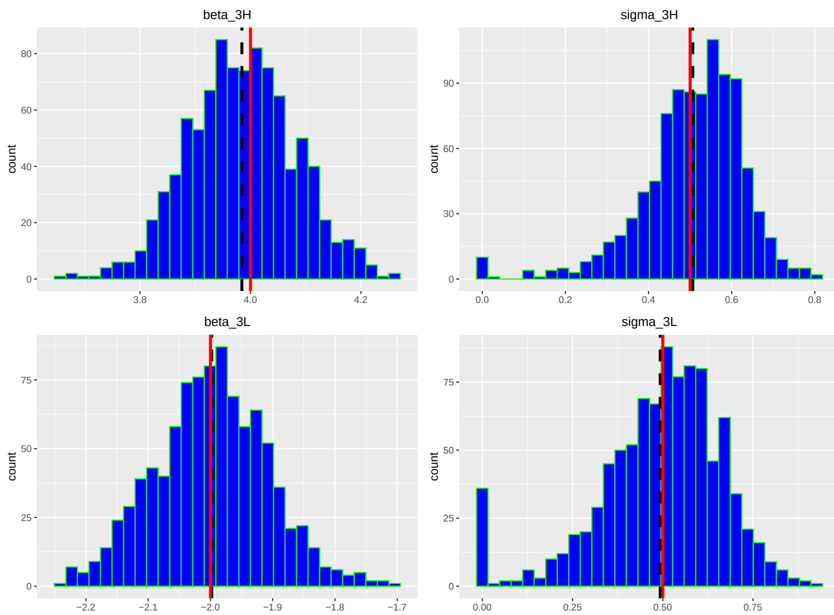


Figure 13: Distribution of estimates for non-linear parameters with “Local interval” instruments ($T = 200, J = 12$)



D.3.1 Estimation with a single Gaussian

Table 18: Estimation with a single Gaussian (1000 replications)

	Instruments	Differentiation						"Optimal"						Interval Global						Interval Local					
	Parameter	β_0	α	β_1	β_2	β_3	σ_3	β_0	α	β_1	β_2	β_3	σ_3	β_0	α	β_1	β_2	β_3	σ_3	β_0	α	β_1	β_2	β_3	σ_3
Sample size	true	2	-2	1.5	1	1.5	0.5	2	-2	1.5	1	1.5	0.5	2	-2	1.5	1	1.5	0.5	2	-2	1.5	1	1.5	0.5
T=50, J=12	bias	-0.16	0.032	-0.031	-0.028	-0.032	-0.004	-0.09	0.018	-0.016	-0.014	-0.018	-0.003	-0.15	0.03	-0.028	-0.026	-0.03	-0.004	-0.15	0.03	-0.028	-0.026	-0.03	-0.001
	$\sqrt{MSE}$	0.292	0.057	0.212	0.209	0.138	0.069	0.27	0.053	0.214	0.211	0.138	0.067	0.288	0.056	0.212	0.209	0.138	0.066	0.286	0.056	0.212	0.209	0.138	0.064
T=50, J=20	bias	-0.091	0.018	-0.022	-0.022	-0.015	0.001	-0.047	0.009	-0.013	-0.013	-0.006	0.001	-0.084	0.017	-0.021	-0.021	-0.013	0	-0.086	0.017	-0.021	-0.021	-0.014	0.002
	$\sqrt{MSE}$	0.209	0.041	0.159	0.16	0.106	0.05	0.199	0.039	0.16	0.161	0.106	0.05	0.206	0.041	0.16	0.16	0.106	0.052	0.208	0.041	0.159	0.16	0.106	0.052
T=100, J=12	bias	-0.088	0.017	-0.001	0	-0.027	0.001	-0.052	0.01	0.007	0.007	-0.02	0.001	-0.082	0.016	0	0.001	-0.026	0.001	-0.074	0.014	-0.016	-0.016	-0.013	0.001
	$\sqrt{MSE}$	0.199	0.039	0.146	0.145	0.1	0.045	0.189	0.037	0.148	0.147	0.099	0.047	0.197	0.039	0.146	0.146	0.1	0.044	0.185	0.036	0.151	0.152	0.099	0.044
T=100, J=20	bias	-0.043	0.009	-0.011	-0.012	-0.006	-0.001	-0.021	0.004	-0.007	-0.008	-0.002	-0.001	-0.04	0.008	-0.011	-0.012	-0.006	-0.001	-0.035	0.007	-0.01	-0.009	-0.004	0
	$\sqrt{MSE}$	0.145	0.028	0.115	0.114	0.075	0.035	0.141	0.028	0.115	0.114	0.075	0.035	0.145	0.028	0.115	0.114	0.076	0.035	0.14	0.027	0.116	0.115	0.076	0.035
T=100, J=20	bias	-0.038	0.007	-0.012	-0.012	-0.004	0.001	-0.017	0.003	-0.006	-0.007	-0.001	0	-0.032	0.006	-0.009	-0.01	-0.004	0	-0.033	0.006	-0.009	-0.01	-0.004	0.001
	$\sqrt{MSE}$	0.132	0.026	0.11	0.11	0.073	0.032	0.127	0.025	0.109	0.109	0.069	0.032	0.129	0.026	0.109	0.109	0.069	0.032	0.129	0.026	0.109	0.109	0.069	0.031

E Empirical application

First stage regression: instruments on price

Table 19: Estimation results - Logit and Nested Logit

	OLS		*instrumental variable*		
	(1)	(2)	(3)	(4)	(5)
Price/income	-0.354^{***}	-2.907^{***}	-2.356^{***}	-2.729^{***}	-2.615^{***}
	(0.041)	(0.133)	(0.124)	(0.053)	(0.052)
log(within market shares)				0.420^{***}	0.407^{***}
				(0.006)	(0.006)
Fuel Cost	-0.210^{***}	-0.138^{***}	-0.247^{***}	-0.074^{***}	-0.126^{***}
	(0.008)	(0.006)	(0.009)	(0.004)	(0.006)
Size(m^2)	0.031	0.001	0.158^{***}	-0.001	0.104^{***}
	(0.038)	(0.040)	(0.041)	(0.025)	(0.026)
Horsepower(KW/100)	0.136	3.151^{***}	2.511^{***}	2.586^{***}	2.431^{***}
	(0.089)	(0.183)	(0.172)	(0.080)	(0.078)
Foreign	0.351^{***}	0.083	0.120^{*}	-0.106^{**}	-0.101^{**}
	(0.064)	(0.073)	(0.070)	(0.046)	(0.044)
Height(m)	0.870^{***}	1.505^{***}	3.487^{***}	1.121^{***}	2.270^{***}
	(0.216)	(0.197)	(0.228)	(0.125)	(0.145)
Gasoline	1.399^{***}	0.625^{***}	1.118^{***}	0.190^{***}	0.422^{***}
	(0.055)	(0.061)	(0.063)	(0.039)	(0.041)
Fuel cost$imes$ income	0.020^{***}	-0.002^{**}	0.014^{***}	-0.002^{***}	0.007^{***}
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
Size $imes$ income	-0.005^{***}	-0.002^{***}	-0.006^{***}	0.0003	-0.002^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Horsepower$imes$ income	0.009^{***}	-0.026^{***}	-0.017^{***}	-0.027^{***}	-0.024^{***}
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Horsepower$imes$ time	-0.084^{***}	-0.068^{***}	-0.083^{***}	-0.038^{***}	-0.045^{***}
	(0.006)	(0.007)	(0.007)	(0.004)	(0.004)
Foreign $imes$ income	-0.019^{***}	-0.015^{***}	-0.016^{***}	-0.008^{***}	-0.008^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Height $imes$ income	-0.006	0.032^{***}	-0.002	0.016^{***}	-0.003
	(0.004)	(0.004)	(0.005)	(0.003)	(0.003)
Height $imes$ density	-0.037^{***}	-0.003^{***}	-0.037^{***}	-0.001^{***}	-0.021^{***}
	(0.004)	(0.0003)	(0.004)	(0.0002)	(0.003)
Gasoline$imes$ income	-0.016^{***}	-0.003^{***}	-0.010^{***}	0.0004	-0.003^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
X2p2015s		-0.024		-0.019	
		(0.019)		(0.012)	
Constant	-7.937^{***}	-12.482^{***}	-11.171^{***}	-9.144^{***}	-8.506^{***}
	(0.167)	(0.149)	(0.167)	(0.092)	(0.102)
State FE/ Year FE	*Yes*	*No*	*Yes*	*No*	*Yes*
Observations	39,888	39,888	39,888	39,888	39,888

Baseline specifications: logit and nested logit

Construction of the interval instruments

- **Discretization of the support**
- **normalization of the instruments**

E.0.1 Results differentiation instruments

Table 20: counterfactual quantities under different specifications on RCs (20 most popular cars)

Counterfactual quantity		Price elasticity			Curvature			Marginal cost			Mark-up			Pass-through		
car	Manufacturer	Logit	Gaussian	Mixture	Logit	Gaussian	Mixture	Logit	Gaussian	Mixture	Logit	Gaussian	Mixture	Logit	Gaussian	Mixture
Golf	Volkswagen	-1.09	-0.95	-3.03	1.00	1.14	1.21	1260	-9670	15436	24098	35028	9922	0.92	-	1.30
Polo	Volkswagen	-0.74	-0.70	-2.50	1.00	1.15	1.09	-6643	-14366	9073	23819	31542	8103	1.05	-	1.09
Passat	Volkswagen	-1.43	-1.21	-2.27	1.00	1.17	1.57	9488	-1033	17826	24631	35153	16294	1.02	-	2.65
Corsa	PSA	-0.66	-0.63	-2.28	1.00	1.14	1.07	-8432	-11246	8410	24088	26902	7246	1.02	-	1.12
Fiesta	Ford	-0.62	-0.60	-2.18	1.00	1.15	1.07	-8983	-10806	7657	23487	25310	6847	1.03	-	1.10
Tiguan	Volkswagen	-1.32	-1.14	-2.28	1.00	1.17	1.55	6831	-2919	16211	24118	33868	14738	1.01	-	2.62
Golf	Volkswagen	-1.17	-1.03	-3.12	1.00	1.18	1.27	3128	-7932	16582	23828	34888	10374	0.99	-	1.41
up!	Volkswagen	-0.53	-0.52	-1.92	1.00	1.14	1.05	-11231	-17703	4594	23278	29749	7453	1.04	-	0.96
Tiguan	Volkswagen	-1.34	-1.15	-3.09	1.00	1.19	1.38	7051	-4117	19186	23842	35009	11706	1.01	-	1.66
1er-Reihe	BMW	-1.16	-1.03	-3.09	1.00	1.18	1.28	3845	-769	19179	25138	29753	9805	0.99	-	1.39
Octavia	Volkswagen	-1.23	-1.08	-2.33	1.00	1.17	1.50	4629	-4504	15464	24211	33345	13377	1.01	-	2.34
A4	Volkswagen	-1.56	-1.30	-2.26	1.00	1.19	1.56	13209	1995	20260	25865	37079	18814	1.01	-	2.66
Clio	Renault	-0.73	-0.70	-2.49	1.00	1.16	1.10	-6240	-8684	9817	23120	25563	7063	1.03	-	1.17
T-Roc	Volkswagen	-0.87	-0.81	-2.80	1.00	1.17	1.14	-3645	-12275	11578	23798	32427	8575	1.06	-	1.16
Kuga	Ford	-1.16	-1.03	-3.09	1.00	1.18	1.28	3654	-518	18214	23684	27856	9124	1.03	-	1.39
Golf	Volkswagen	-1.10	-0.99	-2.34	1.00	1.16	1.44	1548	-7284	13678	23929	32762	11799	0.96	-	2.13
A-Klasse	Daimler	-1.28	-1.10	-3.07	1.00	1.19	1.35	6608	562	20662	25066	31112	11013	1.01	-	1.56
Golf	Volkswagen	-1.05	-0.94	-2.33	1.00	1.16	1.42	417	-8115	13135	24177	32710	11460	0.72	-	2.11
Golf	Volkswagen	-1.18	-1.05	-3.15	1.00	1.18	1.27	3202	-8230	16705	23921	35353	10418	0.98	-	1.40
Octavia	Volkswagen	-1.05	-0.95	-3.02	1.00	1.17	1.21	380	-8835	14808	23862	33077	9433	0.78	-	1.30

Counterfactual quantities under different specifications

Figure 14: Estimated demand functions under different specifications