Identification and Estimation of Demand for Bundles

Alessandro Iaria and Ao Wang*

May 11, 2019

Preliminary and incomplete, please do not circulate without permission

Abstract

We present an instrument-free identification and estimation strategy of a mixed logit model of demand for bundles with endogenous prices given observations for a cross-section of independent markets. Our instrument-free approach hinges on an affine relationship between the utilities of single products and of multi-product bundles typical of this class of models. We achieve identification by relying on an essential real analytic property of the mixed logit model and on a mild restriction on the price-setting model. We also propose a concentrated MLE that substantially alleviates the challenge of dimensionality inherent in estimation. Finally, we apply our methods to estimate demand and supply in the US ready-to-eat cereal industry, where our estimator reduces the numerical search from approximately 20000 to 150 parameters. Our results suggest that ignoring demand synergies among different products often (or rarely) purchased in bundles results in misleading demand estimates and counterfactual simulations.

*University of Bristol (alessandro.iaria@bristol.ac.uk) and CREST (ao.wang@ensae.fr). Iaria gratefully acknowledges funding from the British Academy/Leverhulme SRG no. 170142. Very special thanks to Xavier D'Hautfœuille for the endless discussions and suggestions at each stage of this and related projects. We are extremely grateful for the insightful suggestions to Abi Adams, Gautam Gowrisankaran, Chris Murris, David Pacini, Ariel Pakes, Andrea Pozzi, Pasquale Schiraldi, Howard Smith, Sami Stouli, Ali Yurukoglu, seminar participants at Bristol Internal Workshop, Berlin IO Day, CIREQ Montreal, the 8th EIEF-Unibo-IGIER Bocconi IO Workshop, the International Workshop on Innovation and Industrial Economics (Nanjing), the 29th Jerusalem Summer School of Economics (IO), Guanghua School of Management, PSE, the 17th IIOC, the 2019 GSE Summer Forum Applied IO, the 20th CEPR Applied IO. We would like to thank IRI for making the data used in this paper available. All estimates and analyses in this paper, based on data provided by IRI, are by the authors and not by IRI.

1 Introduction

In standard decision theory, consumer preferences are usually defined over bundles of products rather than over individual products (see Debreu (1959), Varian (1992), and Mas-Colell, Whinston, Green, et al. (1995)), allowing for both substitutability and complementarity among products. Despite some notable exceptions (see Augereau, Greenstein, and Rysman (2006), Gentzkow (2007), and Thomassen, Smith, Seiler, and Schiraldi (2017)), the econometric models routinely used to estimate demand rely on the assumption that each of the products purchased in a bundle is chosen independently from the others, precluding the possibility of complementarity and potentially giving rise to misleading estimates and counterfactual simulations.

Models of demand for bundles face non-trivial identification challenges (see Gentzkow (2007)), even in simple settings with a limited number of products (see Fox and Lazzati (2017)). Moreover, the estimation of demand for bundles is subject to a challenge of dimensionality: the number of parameters can be too large to be handled numerically even with parsimonious specifications (see Berry, Khwaja, Kumar, Musalem, Wilbur, Allenby, Anand, Chintagunta, Hanemann, Jeziorski, et al. (2014)). These challenges forced empirical researchers dealing with demand for bundles either to focus on applications with a very limited number of products (typically two or three) or to make restrictive assumptions on the parameters capturing complementarity and substitutability among the products within bundles (typically assuming a common parameter for all bundles and individuals).

In this paper, we study the identification and estimation of a mixed logit model of demand for bundles with endogenous prices given observations for a cross-section of independent markets.[1] Our arguments hinge on the specific affine relationship between the utilities of individual products and of multi-product bundles typical of models along the lines of Gentzkow (2007)'s: the average utility of any multi-product bundle equals the sum of the average utilities of its individual products plus an extra term capturing the potential demand synergies among the products in the bundle. This utility structure enables us to propose an instrument-free identification approach which is robust to price endogeneity. In addition, it is useful to alleviate the challenge of dimensionality in estimation by means of a novel concept of demand inverse that allows us to concentrate out of the likelihood function the (potentially numerous) market-product specific average utility parameters.

Our approach is based on a (testable) symmetry assumption about the *average* demand synergies across markets: while the demand synergies for any specific bundle may be unobserved and heterogeneous across individuals (accommodating complementarity for some individuals and substitutability for others), their average is required to be identical across markets (but

[1] We focus on the case of cross-sectional data and do not address the case of panel data with repeated observations for each market. The richer information provided by panel data would allow for weaker assumptions than those we make.

potentially different across bundles). Under this symmetry assumption and regularity conditions similar to those by Rothenberg (1971), we derive necessary and sufficient rank conditions for the instrument-free local identification of the model with endogenous prices. This result formalizes Gentzkow (2007)'s insight that, when the average demand synergies are constant across markets, the availability of purchase data for different markets helps identification.

Building on the local identification results, we provide sufficient conditions for the instrument-free global identification of the model with price endogeneity. Our global identification argument combines two main ingredients: an essential property of the mixed logit and a mild restriction on the price-setting model. First, we demonstrate that, given any distribution of random coefficients (parametric or non-parametric), the mixed logit market share function is real analytic with respect to the market-product specific average utility parameters. Second, we assume the existence of exogenous cost shifters that are unobserved but identifiable from observed market shares and prices. Importantly, this assumption does not constrain the observed prices to be the unique equilibrium of the price-setting model, which is allowed to generate several or even infinitely many equilibria.

We propose a concentrated Maximum Likelihood Estimator (MLE) to be implemented with observed bundle-level market shares subject to sampling error and robust to price endogeneity. We account for sampling error to accommodate the typical necessity of computing bundle-level market shares from a sample of household-level purchases (as in Gentzkow (2007), Kwak, Duvvuri, and Russell (2015), Grzybowski and Verboven (2016), Ruiz, Athey, and Blei (2017), and Ershov, Laliberté, and Orr (2018)).[2] The estimation of demand for bundles is subject to a well known challenge of dimensionality: the number of market-product specific average utility parameters and of demand synergy parameters specific to each bundle can be too large to be handled numerically (see Berry, Khwaja, Kumar, Musalem, Wilbur, Allenby, Anand, Chintagunta, Hanemann, Jeziorski, and Mele (2014)). We tackle this practical bottleneck by a novel concept of demand inversion specific to models along the lines of Gentzkow (2007)'s. For any given value of the other model parameters, we establish a one-to-one mapping between the observed product specific *marginal* market shares and the market-product specific average utility parameters.[3] This allows us to concentrate out of the likelihood the potentially very large number of market-product specific average utility parameters and to substantially reduce the dimensionality of the MLE's numerical search. As an example, in our application the MLE's numerical search is reduced from approximately 20000 to 150 parameters. We show that our

[2]Our asymptotics is over the number of individuals in each market, keeping fixed the number of markets and bundles. When the bundle-level market shares are perfectly observed (as in Crawford and Yurukoglu (2012) and Crawford, Lee, Whinston, and Yurukoglu (2018)), the statistical error in our model vanishes and estimation trivially coincides with identification.

[3]Note that our demand inverse differs from the classic one proposed by Berry (1994) and then generalized by Berry, Gandhi, and Haile (2013). For any given value of the other model parameters, Berry (1994) establishes a one-to-one mapping between the observed *bundle* specific market shares and the market-*bundle* specific average utilities.

assumptions for global identification guarantee consistency and asymptotic normality of this concentrated MLE.

We apply our methods to investigate the welfare implications of bundle-by-bundle pricing in the ready-to-eat (RTE) cereal industry in the USA. As we illustrate later, in approximately 20% of their shopping trips, the households in our data purchase two or more different brands of RTE cereals while producers are observed to price their RTE cereals brand-by-brand (i.e., pure components pricing), so that the price of any bundle of brands corresponds to the sum of the prices of each brand. On the one hand, given the substantial purchases of bundles by households, bundle-by-bundle pricing (i.e., mixed bundling pricing) can be more profitable than pure components pricing due to a better ability to price discriminate (see Armstrong (2016)). On the other hand, as noticed by Anderson and Leruth (1993) and Thanassoulis (2007), the possibility to purchase bundles of brands owned by *different* producers introduces a negative externality that can undermine the profit gains of mixed bundling pricing.[4]

Whether mixed bundling pricing would ultimately benefit or harm RTE cereal producers in the USA depends on which of the two mechanisms prevailed given households' preferences and producers' costs. Stressing the importance of accounting for demand synergies, our estimates show that demand for RTE cereals exhibits substantial complementarity among different brands.[5] Given these, our counterfactual simulations highlight that the profit gains of mixed bundling pricing with respect to pure components pricing are sharply decreasing in the level of competition: while a monopolist would benefit from mixed bundling, the observed oligopoly would *not*—even ignoring potential increases in logistics costs. Because of the more effective price discrimination, mixed bundling leads to lower levels of consumer surplus than pure components for any market structure.

Related Literature. In the context of identification, the closest papers to ours are Berry and Haile (2014) and Fox and Lazzati (2017). Berry and Haile (2014) propose sufficient conditions for the non-parametric identification of general demand systems. The mixed logit model we study can be seen as a special case of the general framework by Berry and Haile (2014). In this context, their argument relies on the availability of instruments to pin down the parameters of the distribution of random coefficients (in addition to those necessary to address the problem of price endogeneity). Fox and Lazzati (2017) propose novel sufficient conditions for the non-parametric identification of demand for bundles. Their argument, based on exclusion restrictions in the form of additively separable excluded regressors, naturally applies to the two-

[4]Any increase in the price of an individual brand will directly reduce the demand for all *inter*-producer bundles that include it, so that all producers selling any brand included in these—now more expensive—bundles will be negatively affected. Producers do not fully internalize this and will tend to choose excessively "high" prices for their individual brands. Crawford and Yurukoglu (2012) find empirical evidence of a similar phenomenon in the multichannel television industry in the USA, see their footnote 11.

[5]We consider two brands as complements whenever their cross-price elasticities of demand are negative.

product case, and can be extended to larger choice sets at the cost of additional restrictions.[6] Our identification results complement those by Berry and Haile (2014) and by Fox and Lazzati (2017) in at least three ways. First, our basic identification argument is not based on the availability of instruments or excluded regressors but rather on a symmetry restriction among the average demand synergies across different markets.[7] The alternative arguments may be more or less appropriate in different applications. Second, while Fox and Lazzati (2017) can be preferred in two-product applications with exogenous regressors, our results apply more readily to cases with larger choice sets and endogenous prices, similar to Berry and Haile (2014). Third, our focus on parametric rather than non-parametric models naturally leads to convenient MLEs that alleviate the challenge of dimensionality mentioned earlier.

In the context of estimation, Compiani (2017) proposes a non-parametric estimator of demand models that accommodates a broad range of consumer behaviors, including, among others, demand synergies across products. His estimator builds on the identification argument of Berry and Haile (2014). There is a trade-off between our proposed estimators and Compiani (2017)'s. His non-parametric estimator is more flexible than ours, but it is subject to a severe curse of dimensionality that may constrain its applicability to settings with small choice sets. Differently, our parametric estimators are less affected by dimensionality concerns and can be implemented with larger choice sets.

Our RTE cereal application builds on the classic study by Nevo (2001), and complements it by investigating the welfare consequences of bundle-level pricing of brands that are often (or rarely) observed to be jointly purchased. In terms of bundle-level pricing, the closest empirical paper to ours is Chu, Leslie, and Sorensen (2011), which introduces the concept of bundle-size pricing (i.e., one price for all bundles of a given size) and shows that in the context of theater plays, it can approximate the returns of mixed bundling pricing (98.5%) despite its simplicity. Chu, Leslie, and Sorensen (2011) use data for the plays offered by a single theater company, while we extend the empirical study of bundle-level pricing to an oligopolistic industry and specifically investigate its relationship to market structure. More recently, **?** empirically investigate the classic Cournot (1838)'s question of whether mergers can be welfare enhancing with demand complementarities under pure components pricing (they do not study non-linear pricing). In particular, they study demand for products across different categories, potato chips and carbonated sodas. Differently, we explore the relationship between bundle-level pricing and market structure among producers operating within the same product category, the RTE cereal industry.

[6] The extension to larger choice sets requires further restrictions on the demand synergies, which are not allowed to be heterogeneous across individuals and need to satisfy specific sign constraints (e.g., all products must be assumed to be complements). See the on-line appendix of Fox and Lazzati (2017) for more details.

[7] In more abstract terms, all these alternative identification strategies are indeed based on exclusion restrictions. However, the type of exclusion restriction we propose is of a different nature from those used by the other two papers.

Organization. In the next section, we introduce model and notation. In sections 3 and 4, we present—respectively—our local and global identification results. In section 5, we propose convenient estimators. Finally, in section 6 we conclude the paper with an empirical application.

2 Model and Notation

Let there be a cross-section of $t = 1, ..., T$ independent markets, each populated by $i = 1, ..., I$ individuals. Individual i in market t exclusively makes purchases in market t and is a different person from individual i in any other market t'. For individuals in market t, let $\mathbf{J}_t$ be the set of $j = 1, ..., J_t$ market-specific products that can be purchased in isolation or in bundles. Let $\mathbf{C}_t$ be the market t-specific choice set of bundles available for purchase. Denote the set of non-empty bundles in market t by $\mathbf{C}_{t1}$, the full choice set by $\mathbf{C}_t = \mathbf{C}_{t1} \cup \{0\}$, where 0 is the empty bundle (i.e., the outside option), and the set of *multi-product* bundles in market t by $\mathbf{C}_{t2} \subseteq \mathbf{C}_{t1}$. This set does not include the *single-product* bundles: those bundles made of just a single product $j \in \mathbf{J}_t$. The indirect utility of individual i in market t from choosing product j is:

$$\begin{aligned} U_{itj} &= u_{itj} + \varepsilon_{itj} \\ &= \delta_{tj} + \mu_{itj} + \varepsilon_{itj} \quad \text{and} \\ U_{it0} &= \varepsilon_{it0}, \end{aligned} \tag{1}$$

where $u_{itj} = \delta_{tj} + \mu_{itj}$, δ_{tj} is market t-specific average utility for product $j \in \mathbf{J}_t$, μ_{itj} is an individual i-specific utility deviation from δ_{tj}, while ε_{itj} and ε_{it0} are error terms. Throughout the paper, we will treat the market t-specific average utilities as parameters to be identified and estimated. However, one can reduce the number of parameters by employing observable characteristics and making additional functional form assumptions.[8]

To ease exposition, we refer to the products in bundle $\mathbf{b}$ as to $j \in \mathbf{b}$. Following Gentzkow (2007), the indirect utility of individual i in market t from choosing bundle $\mathbf{b} \in \mathbf{C}_{t2}$ is:

$$\begin{aligned} U_{it\mathbf{b}} &= \sum_{j \in \mathbf{b}} u_{itj} + \Gamma_{it\mathbf{b}} + \varepsilon_{it\mathbf{b}} \\ &= \sum_{j \in \mathbf{b}} (\delta_{tj} + \mu_{itj}) + \Gamma_{t\mathbf{b}} + (\Gamma_{it\mathbf{b}} - \Gamma_{t\mathbf{b}}) + \varepsilon_{it\mathbf{b}} \\ &= \sum_{j \in \mathbf{b}} \delta_{tj} + \Gamma_{t\mathbf{b}} + \left[\sum_{j \in \mathbf{b}} \mu_{itj} + \zeta_{it\mathbf{b}} \right] + \varepsilon_{it\mathbf{b}} \\ &= \delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}) + \mu_{it\mathbf{b}} + \varepsilon_{it\mathbf{b}}, \end{aligned} \tag{2}$$

where $\Gamma_{it\mathbf{b}}$ is the individual-market it-specific *demand synergy* among the products of bundle

[8] We provide more detail on this while discussing price endogeneity at the end of this section.

$\mathbf{b}$, which we specify as $\Gamma_{it\mathbf{b}} = \Gamma_{t\mathbf{b}} + \zeta_{it\mathbf{b}}$. $\Gamma_{t\mathbf{b}}$ is the average demand synergy for the products in bundle $\mathbf{b}$ among the individuals in market t and $\zeta_{it\mathbf{b}}$ is an individual-market it-specific deviation from this average. $\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}) = \sum_{j\in\mathbf{b}} \delta_{tj} + \Gamma_{t\mathbf{b}}$ is market t-specific average utility for bundle $\mathbf{b}$, $\mu_{it\mathbf{b}}$ is an individual i-specific utility deviation from $\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}})$, and $\varepsilon_{it\mathbf{b}}$ is an error term.

The demand synergy parameter $\Gamma_{it\mathbf{b}}$ captures the extra utility individual i in market t obtains from buying the products in bundle $\mathbf{b}$ jointly rather than separately. When $\Gamma_{it\mathbf{b}} > 0$, the utility of the bundle is super-modular with respect to the utilities of the single products and, from i's perspective, the products are complements. Conversely, when $\Gamma_{it\mathbf{b}} < 0$, from i's perspective the products in the bundle are substitutes.[9] Similar to the t-specific average utilities δ_{tj}, we treat also the t-specific average demand synergies $\Gamma_{t\mathbf{b}}$ as parameters to be identified and estimated. However, as discussed in more detail below, in applications with observable bundle-specific characteristics (e.g., bundle-specific discounts), one can make further functional form assumptions and specify $\Gamma_{t\mathbf{b}}$ in terms of these characteristics.

We now turn to our distributional assumptions for the unobserved components of utility: μ_{itj} for each $j \in \mathbf{J}_t$, $\zeta_{it\mathbf{b}}$ and ε_{itj} for each $\mathbf{b} \in \mathbf{C}_t$. We assume that $\mu_{it\mathbf{b}} = \sum_{j\in\mathbf{b}} \mu_{itj} + \zeta_{it\mathbf{b}}$ can be expressed as a function of a vector of random coefficients β_{it}, so that $\mu_{it\mathbf{b}} = \mu_{it\mathbf{b}}(\beta_{it})$, and that β_{it} is distributed according to $F(\cdot;\Sigma)$, where Σ is a finite-dimensional parameter in some connected compact set $\Theta_\Sigma \subset \mathbb{R}^P$. As is typical in empirical papers, the random coefficients β_{it} can be specified so that $\mu_{it\mathbf{b}}(\beta_{it})$ be a function of observable demographics and/or product and bundle characteristics. Moreover, $\varepsilon_{it\mathbf{b}}$ for $\mathbf{b} \in \mathbf{C}_t$ is assumed to be i.i.d. Gumbel.

Even though we make the assumption that $\varepsilon_{it\mathbf{b}}$ is i.i.d. Gumbel, the distribution of the random coefficients β_{it}, $F(\cdot;\Sigma)$, is allowed to belong to any known parametric family. As shown by McFadden and Train (2000), under mild regularity conditions, any discrete choice model derived from random utility maximization can be approximated arbitrarily well by mixed logit models of the kind we consider. In addition, note that our mixed logit model is a generalization of Gentzkow (2007)'s, which restricts $F(\cdot;\Sigma)$ to be a normal distribution and $\Gamma_{it\mathbf{b}} = \Gamma_{\mathbf{b}}$ for all i's and t's. We add a layer of unobserved heterogeneity to the individual preferences specific to each bundle: for reasons unobserved to the econometrician, the products in any bundle can be seen as complements by some individuals and as substitutes by others.

Denote market t-specific average utility vector by $\delta_t(\Gamma_t) = (\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}))_{\mathbf{b}\in\mathbf{C}_{t1}}$ and the vector collecting all the market t-specific average demand synergies by Γ_t. $\delta_t(\Gamma_t)$ does not only list the t-specific average utilities of multi-product bundles $\mathbf{b} \in \mathbf{C}_{t2}$, but also those of the single-product bundles $\mathbf{b} \in \mathbf{C}_{t1} \setminus \mathbf{C}_{t2}$: given that single-product bundles have zero demand synergies, our notation for any $\mathbf{b} = j \in \mathbf{C}_{t1} \setminus \mathbf{C}_{t2}$ is $\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}) = \delta_{tj}$. Given our distributional assumptions, the market share function of bundle $\mathbf{b} \in \mathbf{C}_t$ for individuals in market t takes the mixed logit

[9]For an axiomatic characterization of the concepts of complementarity and substitutability, see Manzini, Mariotti, and Ülkü (2018).

form:

$$
\begin{aligned}
s_{t\mathbf{b}}(\delta_t(\Gamma_t);\Sigma) &= \int s_{it\mathbf{b}}(\delta_t(\Gamma_t),\beta_{it})dF(\beta_{it};\Sigma) \\
&= \int \frac{e^{\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}})+\mu_{it\mathbf{b}}(\beta_{it})}}{\sum_{\mathbf{b}'\in\mathbf{C}_t} e^{\delta_{t\mathbf{b}'}(\Gamma_{t\mathbf{b}'})+\mu_{it\mathbf{b}'}(\beta_{it})}} dF(\beta_{it};\Sigma),
\end{aligned}
\tag{3}
$$

where $s_{it\mathbf{b}}(\delta_t(\Gamma_t),\beta_{it})$ is individual i's choice probability of bundle $\mathbf{b}$ in market t given β_{it}.

Interpretation of Demand Synergies. We follow Gentzkow (2007)'s narrative for the interpretation of the demand synergy parameter $\Gamma_{it\mathbf{b}}$ as capturing complementarity or substitutability in consumption. However, demand for bundles—and consequently demand synergies—can also arise for different reasons: for example, shopping costs (as in Thomassen, Smith, Seiler, and Schiraldi (2017)) and preference for variety (as in Hendel (1999) and Dubé (2004)). If individuals face shopping costs every time they visit a store, they can find more convenient to purchase all their products at once rather than over several trips (one-stop shoppers). In our model, shopping costs can be rationalized by positive demand synergies. In Appendix 7.1 we also illustrate that the model of preference for variety by Hendel (1999) and Dubé (2004) is a special case of ours with negative demand synergies. The demand synergy $\Gamma_{it\mathbf{b}}$ is a catch-all parameter that can reflect complementarity or substitutability in consumption as in Gentzkow (2007), shopping costs as in Thomassen, Smith, Seiler, and Schiraldi (2017), or preference for variety as in Hendel (1999) and Dubé (2004).

Random Intercepts and Demand Synergies. As argued by Gentzkow (2007), the random intercepts $(\mu_{itj})_{j=1}^{J_t}$ play an important conceptual role in the identification of demand synergies in mixed logit models of demand for bundles. Without random coefficients, the Independence from Irrelevant Alternatives (IIA) property would imply that the relative predicted market shares of any two bundles do not depend on the characteristics of any other bundle. Removing from the choice set a bundle almost identical to the preferred one (e.g., same products but one) or a bundle completely different from it (e.g., only different products) would equivalently have no impact on the remaining relative predicted market shares. The random intercepts mitigate this limitation in an intuitive way: the indirect utilities of all bundles including product j will share the random intercept μ_{itj}, so that bundles with a larger overlap of products will also have more correlated indirect utilities. Ultimately, not accounting for possible correlations across the indirect utilities of bundles with overlapping products may lead to finding spurious demand synergies (see Gentzkow (2007)).

Average Utilities and Price Endogeneity. As mentioned above, we treat the average utility δ_{tj} as a fixed effect to be identified and estimated, being vague about its exact dependence on prices and other observed or unobserved market-product specific characteristics. For example,

following Berry (1994) and BLP, a classical linear specification is $\delta_{tj} = x_{tj}\alpha_x + \alpha_p p_{tj} + \xi_{tj}$, where x_{tj} is a vector of exogenous observed characteristics, p_{tj} is the observed price, $\alpha = (\alpha_x, \alpha_p)^{\mathrm{T}}$ is a vector of preference parameters, and ξ_{tj} is a residual unobserved by the econometrician but observed by both individuals and price-setting firms. In this context, endogeneity arises whenever prices are chosen by firms on the basis of $(\xi_{tj})_{j=1}^{J_t}$.

Our local identification arguments are robust to cases of price endogeneity in which, for any bundle $\mathbf{b}$, the source of endogeneity is confined to $\delta_{t\mathbf{b}}(\Gamma_{\mathbf{b}}) = \sum_{j\in\mathbf{b}} \delta_{tj} + \Gamma_{\mathbf{b}}$ (see Assumption 2 below). While this allows δ_{tj} to be any arbitrary function of $(x_{tj}, p_{tj}, \xi_{tj})$, it is not compatible with bundles being priced on the basis of market-bundle specific unobservables (e.g., $\xi_{t\mathbf{b}}$). Our global identification arguments require further restrictions both on δ_{tj} and on the price-setting model. As detailed in section 4.2, we require: (i) the average utility δ_{tj} to be additively separable in ξ_{tj} and an arbitrary function of (x_{tj}, p_{tj}) and (ii) the existence of exogenous cost shifters that are unobserved by the econometrician but identifiable from observed market shares and prices. Importantly, (ii) does not restrict the observed prices to be the unique equilibrium of the price-setting model, which can generate several or even infinitely many equilibria.

3 Local Identification

Suppose that the econometrician observes without error the market shares $\mathscr{s}_{t\mathbf{b}}$ of each bundle $\mathbf{b} \in \mathbf{C}_{t1}$ for each independent market $t = 1, ..., T$.[10,11] We focus on the case of a fixed number of products J_t and of a fixed number of independent markets T. We do not consider the case of panel data with repeated observations for each market.[12] Similar to Berry and Haile (2014), our notion of identification concerns the conditions under which

$$\begin{aligned} & s_{t\mathbf{b}}(\delta'_t(\Gamma'_t); \Sigma') = \mathscr{s}_{t\mathbf{b}} \\ & \text{subject to } \Gamma'_{t\mathbf{b}} = \delta'_{t\mathbf{b}}(\Gamma'_{t\mathbf{b}}) - \sum_{j\in\mathbf{b}} \delta'_{tj}, \mathbf{b} \in \mathbf{C}_{t2} \end{aligned} \tag{4}$$

has a unique solution for $t = 1, ..., T$ and $\mathbf{b} \in \mathbf{C}_{t1}$, where $\delta'_t(\Gamma'_t) = (\delta'_{t\mathbf{b}}(\Gamma'_{t\mathbf{b}}))_{\mathbf{b}\in\mathbf{C}_{t1}}$ and $s_{t\mathbf{b}}(\delta'_t(\Gamma'_t); \Sigma')$ is defined in (3).[13] Define the $J_t \times 1$ market t-specific vector $\delta_{t\mathbf{J}_t} = (\delta_{tj})_{j\in\mathbf{J}_t}$, and the $C_{t1} \times 1$ market t-specific vectors $s_t(\cdot; \Sigma') = \big(s_{t\mathbf{b}}(\cdot; \Sigma')\big)_{\mathbf{b}\in\mathbf{C}_{t1}}$ and $\mathscr{s}_t = (\mathscr{s}_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_{t1}}$.

[10] This is only for the purpose of identification, in estimation we consider the possibility that observed market shares are subject to sampling error.

[11] Sher and Kim (2014) and Allen and Rehbeck (forthcoming) study a different identification problem, where only the *marginal* market shares of products, rather than the market shares of bundles, are observed.

[12] In the case of panel data, because of the richer information available, identification can be achieved relying on weaker assumption than those we make.

[13] Remember that $\delta_t(\Gamma_t)$ also includes the t-specific average utilities of the single-product bundles ($\mathbf{b} \in \mathbf{C}_{t1} \setminus \mathbf{C}_{t2}$): implicitly, our notation for any single-product bundle $\mathbf{b} = j \in \mathbf{C}_{t1} \setminus \mathbf{C}_{t2}$ is $\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}) = \delta_{tj}$.

Definition 1. *Model (3) is locally identified if and only if there exists a neighborhood V of the true parameters $(\delta_{1\mathbf{J}_1}, ..., \delta_{T\mathbf{J}_T}, \Gamma_1, ..., \Gamma_T, \Sigma)$ such that $(\delta_{1\mathbf{J}_1}, ..., \delta_{T\mathbf{J}_T}, \Gamma_1, ..., \Gamma_T, \Sigma)$ is the unique solution to (4) in V.*

Remark 1. *Definition 1 constrains our discussion of identification to the existence of a unique solution to system (4) in mixed logit model (3). As a consequence, we will refer to the existence of multiple solutions to this specific problem as to lack of identification.*[14] *Because of the non-linear nature of model (3), we start by studying the problem of local identification. In section 4, we then investigate the problem of global identification, which requires stronger assumptions.*

Building on Berry, Gandhi, and Haile (2013), our identification arguments rely on demand inverses derived from (4). Define the inverse market share for market $t = 1, ..., T$ as:

$$s_t^{-1}(\cdot;\Sigma) = \left(s_{t\mathbf{b}}^{-1}(\cdot;\Sigma)\right)_{\mathbf{b}\in\mathbf{C}_{t1}} : \mathbf{S}_{t1} \rightrightarrows \mathbb{R}^{C_{t1}}, \tag{5}$$

where $s_{t\mathbf{b}}^{-1}(\cdot;\Sigma)$ is the inverse market share for market $t = 1, ..., T$ and bundle $\mathbf{b} \in \mathbf{C}_{t1}$, C_{t1} is the number of bundles in $\mathbf{C}_{t1}$, and

$$\mathbf{S}_{t1} = \{(\mathscr{s}_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_{t1}} : \mathscr{s}_{t\mathbf{b}} \in (0,1), \sum_{\mathbf{b}\in\mathbf{C}_{t1}} \mathscr{s}_{t\mathbf{b}} < 1\}$$

is the set of all feasible market share vectors of non-empty bundles for market $t = 1, ..., T$. We rely on the next Assumption throughout the paper, excluded the first statement of Lemma 1 and Theorem 3.

Assumption 1.

1. *$\Theta_\Sigma \subset \Upsilon$ is a compact set, where $\Upsilon \subset \mathbb{R}^P$ is a topological space of $\mathbb{R}^P$.*
2. *The density of β_{it}, $\frac{dF(\beta_{it};\Sigma')}{d\beta_{it}}$, is continuously differentiable with respect to Σ' for any β_{it}.*

Lemma 1.

- *For any given $\Sigma' \in \Theta_\Sigma$, the inverse market share (5) is a well defined* function*: for each $\mathscr{s}_t \in \mathbf{S}_{t1}$, there exists a unique $\delta'_t \in \mathbb{R}^{C_{t1}}$ such that $s_t(\delta'_t;\Sigma') = \mathscr{s}_t$.*
- *Given Assumption 1, the inverse market share, $s_t^{-1}(\mathscr{s}'_t;\Sigma')$, is continuously differentiable with respect to $(\mathscr{s}'_t, \Sigma')$ in a neighbourhood of $(\mathscr{s}_t, \Sigma)$.*

Proof. See Appendix 7.3. □

[14]This is what Sargan (1983) calls *first order* lack of identification.

In Appendix 7.2, we illustrate that even simple versions of model (3) raise non-trivial identification issues. First, we show that without further restrictions on Γ_t or additional external information, model (3) can hardly be identified. Second, we discuss three examples that highlight Gentzkow (2007)'s insight: when $\Gamma_t = \Gamma$, the availability of purchase data for different markets helps identification. In what follows, we study identification under this restriction.

Assumption 2. *$\Gamma_{it\mathbf{b}} = \Gamma_{t\mathbf{b}} + \zeta_{it\mathbf{b}}$, with $\Gamma_{t\mathbf{b}} = \Gamma_{\mathbf{b}}$ for $t = 1, ..., T$ and $\mathbf{b} \in \mathbf{C}_{t2}$, and $\zeta_{it\mathbf{b}}$ representing an unobserved idiosyncratic demand synergy.*

Note that, even though Assumption 2 requires all markets to have the same average $\Gamma_{\mathbf{b}}$ (similar to Gentzkow (2007)), each individual is allowed to have unobserved idiosyncratic demand synergy $\zeta_{it\mathbf{b}}$, so that each *it* may have a $\Gamma_{it\mathbf{b}}$ potentially different from $\Gamma_{\mathbf{b}}$. Assumption 2 is compatible with parametric specifications that allow $\Gamma_{it\mathbf{b}}$ to differ across individuals, markets, and bundles also for observable reasons. For example, if one observed bundle-specific characteristics $x_{t\mathbf{b}}$, such as discounts for certain bundles (e.g., non-linear prices due to purchases of large quantities), one could have $\Gamma_{it\mathbf{b}} = x_{t\mathbf{b}}\gamma + \zeta_{it\mathbf{b}}$. More broadly, the key restriction we require is that the *parameters* involved in the specification of $\Gamma_{it\mathbf{b}}$ only vary at the bundle level, but not at the market level.

Assumption 2 represents an alternative to the availability of instruments for the identification of mixed logit models along the lines of Berry and Haile (2014). The presence of random coefficients, i.e. $\dim(\Sigma) > 0$, leads system (4) to have more unknowns than equations, introducing an identification problem not present in multinomial logit models. In general demand systems where the indirect utilities of different alternatives have no particular relationships, this dimensionality issue is typically addressed by adding external information in the form of instruments (in addition to those necessary to address the problem of price endogeneity). However, in the case of Gentzkow (2007)'s demand for bundles, the specific structure that links the indirect utilities of multi-product bundles to those of individual-product bundles allows to reduce dimensionality from *within* the system. Assumption 2 embodies this strategy: by imposing a symmetry restriction among the average demand synergies across markets, the model can be identified without requiring additional instruments.

Importantly, Assumption 2 gives rise to testable implications. As we will see below, in applications with either a large number of markets T, or with large overlapping choice sets across markets, or both, Assumption 2 leads to potentially many over-identifying moment restrictions. In these cases, one can perform likelihood-ratio tests to challenge various aspects of Assumption 2. For example, one could test whether $\Gamma_{t\mathbf{b}} = \Gamma_{1\mathbf{b}}$ for $t = 1, ..., T_1$, $\Gamma_{t\mathbf{b}} = \Gamma_{2\mathbf{b}}$ for $t = T_1 + 1, ..., T_2$, and so on until each t belongs to one of Q nests of "similar" markets with $\Gamma_{1\mathbf{b}} \neq \Gamma_{2\mathbf{b}} \neq ... \neq \Gamma_{Q\mathbf{b}}$; or whether the vector of observables $x_{t\mathbf{b}}$ in $\Gamma_{it\mathbf{b}} = x_{t\mathbf{b}}\gamma + \zeta_{it\mathbf{b}}$ appears to be correctly specified.

Due to Lemma 1 and Assumption 2, at the true parameters Σ and market shares $\mathscr{s}_t$, one

can re-express the first line of system (4) as:

$$\begin{aligned} &\delta_{t\mathbf{b}}(\Gamma_{\mathbf{b}}) = s^{-1}_{t\mathbf{b}}(\jmath_t;\Sigma), \text{ for multi-product bundle } \mathbf{b} \in \mathbf{C}_{t2} \\ &\delta_{tj} = s^{-1}_{tj}(\jmath_t;\Sigma), \text{ for single-product bundle } j \in \mathbf{b}. \end{aligned} \tag{6}$$

By substituting (6) into the second line of (4), one gets:

$$\Gamma_{\mathbf{b}} = s^{-1}_{t\mathbf{b}}(\jmath_t;\Sigma) - \sum_{j\in\mathbf{b}} s^{-1}_{tj}(\jmath_t;\Sigma), \tag{7}$$

for $t = 1, ..., T$ and multi-product bundle $\mathbf{b} \in \mathbf{C}_{t2}$. Note that the left-hand side of system (7) does not depend on market $t = 1, ..., T$, while the right-hand side does. Consequently, at the true parameters Σ, true market shares of any two markets, $\jmath_t$ and $\jmath_{t'}$, and any $\mathbf{b} \in \mathbf{C}_{t2} \cap \mathbf{C}_{t'2}$, one obtains:

$$s^{-1}_{t\mathbf{b}}(\jmath_t;\Sigma) - \sum_{j\in\mathbf{b}} s^{-1}_{tj}(\jmath_t;\Sigma) = s^{-1}_{t'\mathbf{b}}(\jmath_{t'};\Sigma) - \sum_{j\in\mathbf{b}} s^{-1}_{t'j}(\jmath_{t'};\Sigma). \tag{8}$$

The key intuition to our identification strategy is to explore all such moment conditions for any pair of markets $t \neq t'$ and any $\mathbf{b} \in \mathbf{C}_{t2} \cap \mathbf{C}_{t'2}$. As we will see in Lemma 4, under certain conditions, these moment restrictions can uniquely determine the true parameters Σ. Due to (7), the true parameters Σ can then uniquely determine the true demand synergies $\Gamma_{\mathbf{b}}$, $\mathbf{b} \in \mathbf{C}_{t2}$. Finally, because of Lemma 1, one can uniquely recover $\delta_t(\Gamma) = (\delta_{t1}, ..., \delta_{tJ_t}, (\delta_{t\mathbf{b}}(\Gamma_{\mathbf{b}}))_{\mathbf{b}\in\mathbf{C}_{t2}})$ and therefore the true value of δ_{tj} for $t = 1, ..., T$ and $j = 1, ..., J_t$.[15]

Denote the set of all multi-product bundles by $\mathbf{C}_2 = \cup_{t=1,...,T}\mathbf{C}_{t2}$. Note that for any $\mathbf{b} \in \mathbf{C}_2$, there exists t such that $\mathbf{b} \in \mathbf{C}_{t2}$. Then, for any $\mathbf{b} \in \mathbf{C}_2$, define $\mathbf{T_b} = \{t : \mathbf{b} \in \mathbf{C}_{t2}\}$. If $\mathbf{T_b}$ has more than one element, we order them increasingly from t_1 to $t_{|\mathbf{T_b}|}$, then by applying the right-hand side of (7) to t_a and to t_{a+1} and by taking the difference, for $a = 1, ..., |\mathbf{T_b}| - 1$, we obtain $|\mathbf{T_b}| - 1$ moment conditions:

$$\begin{aligned} m_{\mathbf{b}}(\Sigma';\jmath) = &\Bigg[s^{-1}_{t_a\mathbf{b}}(\jmath_{t_a};\Sigma') - \sum_{j\in\mathbf{b}} s^{-1}_{t_aj}(\jmath_{t_a};\Sigma') \\ &- \Big(s^{-1}_{t_{a+1}\mathbf{b}}(\jmath_{t_{a+1}};\Sigma') - \sum_{j\in\mathbf{b}} s^{-1}_{t_{a+1}j}(\jmath_{t_{a+1}};\Sigma') \Big) \Bigg]_{a=1,...,|\mathbf{T_b}|-1}, \\ m_{\mathbf{b}}(\Sigma';\jmath)|_{\Sigma'=\Sigma} = &\ 0. \end{aligned} \tag{9}$$

Moment conditions (9) rely on relationship (7) and the fact that markets t_a and t_{a+1} *average* demand synergies both equal $\Gamma_{\mathbf{b}}$. This implies that, at the true parameter values $\Sigma' = \Sigma$, $m_{\mathbf{b}}(\Sigma';\jmath)|_{\Sigma'=\Sigma} = (\Gamma_{\mathbf{b}} - \Gamma_{\mathbf{b}})_{a=1,...,|\mathbf{T_b}|-1} = 0$. Define $m(\Sigma') = m(\Sigma';\jmath)$ as a function of $\Sigma' \in \Theta_\Sigma$

[15]Remember that $j = 1, ..., J_t$ denotes the single-product bundles available to individuals in market t.

that stacks together the above moment conditions for all the multi-product bundles $\mathbf{b}$ with $|\mathbf{T_b}| \geq 2$: $m(\Sigma') = (m_{\mathbf{b}}(\Sigma'; \jmath))_{\mathbf{b}\in\mathbf{C}_2, |\mathbf{T_b}|\geq 2}$. We then have $m(\Sigma')|_{\Sigma'=\Sigma} = 0$, which consists of $\sum_{\mathbf{b}\in\mathbf{C}_2,|\mathbf{T_b}|\geq 2}(|\mathbf{T_b}| - 1)$ moment conditions with $P = \dim(\Sigma')$ unknowns.

Theorem 1. *If Assumptions 1 and 2 hold, and the Jacobian matrix* $\frac{\partial m(\Sigma')}{\partial \Sigma'}\big|_{\Sigma'=\Sigma}$ *is of full column rank, then model (3) is locally identified.*

Proof. See Appendix 7.4. □

Note that this full column rank condition is the non-linear analogue of the classic *no perfect multicollinearity* condition for the identification of the linear regression model.[16] In what follows, inspired by Rothenberg (1971), we show that our proposed full column rank condition is also necessary for identification among the *rank regular* $\Sigma \in \Theta_\Sigma$.[17,18] Rank regularity is a broader concept than full column rank: if $\frac{\partial m(\Sigma')}{\partial \Sigma'}\big|_{\Sigma'=\Sigma}$ is of full column rank, i.e. $\text{rank}(\frac{\partial m(\Sigma')}{\partial \Sigma'})\big|_{\Sigma'=\Sigma} = \dim(\Sigma) = P$, then Σ is rank regular.[19]

Theorem 2. *If Assumptions 1 and 2 hold, and* $\Sigma \in \Theta_\Sigma$ *is rank regular for* $\frac{\partial m(\Sigma')}{\partial \Sigma'}$. *Then, model (3) is locally identified if and only if* $\frac{\partial m(\Sigma')}{\partial \Sigma'}\big|_{\Sigma'=\Sigma}$ *is of full column rank.*

Proof. See Appendix 7.6 □

Theorem 2 establishes the link between the number of markets with *overlapping* choice sets and the local identification of model (3). Note that, if the number of markets with bundle $\mathbf{b}$ available in the choice set increases, so that $|\mathbf{T_b}|$ becomes larger, then the number of moment conditions in (9) increases. In this sense, Theorem 2 formalizes the intuition that having data on additional markets with overlapping choice sets, or analogously on larger overlapping choice sets for certain markets, will help identification by increasing the number of moment conditions. Specifically, suppose that Σ is rank regular and that its dimension, P, is greater than the number of moment conditions, $\sum_{\mathbf{b}\in\mathbf{C}_2,|\mathbf{T_b}|\geq 2}(|\mathbf{T_b}| - 1)$. Then, the rank of $\frac{\partial m(\Sigma')}{\partial \Sigma'}\big|_{\Sigma'=\Sigma}$ cannot exceed the number of its rows, $\sum_{\mathbf{b}\in\mathbf{C}_2,|\mathbf{T_b}|\geq 2}(|\mathbf{T_b}| - 1)$, which in turn is smaller than

[16] For a general discussion on full column rank conditions and identification in non-linear models, see Newey and McFadden (1994) and Lewbel (forthcoming).

[17] $\Sigma \in \Theta_\Sigma$ is rank regular for the continuously differentiable function $\frac{\partial m(\Sigma')}{\partial \Sigma'}$ if there exists a neighbourhood U of Σ such that $\text{rank}(\frac{\partial m(\Sigma')}{\partial \Sigma'}) = \text{rank}(\frac{\partial m(\Sigma')}{\partial \Sigma'})\big|_{\Sigma'=\Sigma}$ for each $\Sigma' \in U$.

[18] Rothenberg (1971) shows the usefulness of the concept of rank regularity for local identification in non-linear models. Our Theorem 2 adapts Rothenberg (1971)'s Theorem 1 (p. 579) to our environment. Note that the concept of rank regularity is not vacuous in our context and there is plenty of such points: the set of rank regular points of $\frac{\partial m(\Sigma')}{\partial \Sigma'}$ is open and dense in Θ_Σ. For a proof of this property, see Appendix 7.5.

[19] In fact, $\left[\frac{\partial m(\Sigma')}{\partial \Sigma'}\right]^T \left[\frac{\partial m(\Sigma')}{\partial \Sigma'}\right]$ has positive determinant at $\Sigma' = \Sigma$. Moreover, $\frac{\partial m(\Sigma')}{\partial \Sigma'}$ is continuously differentiable with respect to Σ'. Then, the determinant of $\left[\frac{\partial m(\Sigma')}{\partial \Sigma'}\right]^T \left[\frac{\partial m(\Sigma')}{\partial \Sigma'}\right]$ is also continuous with respect to Σ' and therefore positive in a neighbourhood of $\Sigma' = \Sigma$. As a consequence, $\frac{\partial m(\Sigma')}{\partial \Sigma'}$ is of full column rank in a neighbourhood of $\Sigma' = \Sigma$ and has constant rank P in the same neighbourhood of $\Sigma' = \Sigma$.

the number of its columns, P. As a consequence, $\frac{\partial m(\Sigma')}{\partial \Sigma'}\big|_{\Sigma'=\Sigma}$ is not of full column rank and model (3) is not identified.

As mentioned above, in addition to helping with just-identification, the availability of rich data on several markets with large overlapping choice sets will often enable one to attain over-identification, i.e. $\sum_{\mathbf{b}\in\mathbf{C}_2,|\mathbf{T_b}|\geq 2}(|\mathbf{T_b}|-1) > P$, so to perform likelihood-ratios tests and to challenge various features of Assumption 2.

While theoretically useful, the concept of rank regularity is abstract and not easily verifiable in practice. However, there are simpler and practically more appealing conditions that imply rank regularity and the applicability of Theorem 2. The next Corollary shows that whenever the dimension of Σ is larger than the number of moment conditions and the Jacobian matrix $\frac{\partial m(\Sigma')}{\partial \Sigma'}\big|_{\Sigma'=\Sigma}$ is of full *row* rank, then Σ is rank regular and model (3) is *not* identified.[20]

Corollary 1. *Suppose Assumptions 1 and 2 hold, and the number of moment conditions,* $\sum_{\mathbf{b}\in\mathbf{C}_2,|\mathbf{T_b}|\geq 2}(|\mathbf{T_b}|-1)$ *is strictly smaller than the dimension of* Σ, P. *Then, if the Jacobian matrix* $\frac{\partial m(\Sigma')}{\partial \Sigma'}\big|_{\Sigma'=\Sigma}$ *is of full row rank, model (3) is not locally identified.*

Proof. See Appendix 7.7. □

4 Global Identification

Our identification discussion so far focused on the local uniqueness of solutions to system (9). Without any further restriction, the set of solutions to system (9) over the entire domain of parameters may not be singleton. There are at least two approaches to dealing with this global multiplicity of solutions. Partial identification, which entails the characterization of the set of global solutions to system (9), i.e. the identified set, and global identification, which consists in strengthening the conditions for local identification until the identified set is singleton over the entire domain of parameters. We opt for the second approach and, in what follows, discuss sufficient conditions for global identification. Our choice is pragmatic and motivated by estimation convenience: as detailed in section 5, our global identification conditions imply consistency and asymptotic normality of a convenient concentrated MLE.

In the first part of our argument, we extend the full column rank condition from $\Sigma' = \Sigma$ to the entire domain Θ_Σ.[21] We denote by $m(\Sigma'; \mathbf{T}_0)$ moment conditions (9) constructed from markets in $\mathbf{T}_0 \subsetneq \mathbf{T}$ and evaluated at Σ'.

Assumption 3. *There exists* $\mathbf{T}_0 \subsetneq \mathbf{T}$ *such that* $\frac{\partial m(\Sigma';\mathbf{T}_0)}{\partial \Sigma'}$ *is of full column rank for every* $\Sigma' \in \Theta_\Sigma$.

[20]Note that lack of local identification is the strongest negative result one can get: if the model is not locally identified, then for sure it will not be globally identified.

[21]This condition is standard in the global identification literature. See for example Gale and Nikaido (1965) and Berry, Gandhi, and Haile (2013). A notable exception is Komunjer (2012).

Lemma 2. *If Assumptions 1–3 hold, and Θ_Σ is compact, then the number of solutions to system (9) in Θ_Σ is finite.*

Proof. See Appendix 7.8. $\square$

The second part of our argument hinges on an essential property of the mixed logit model: given *any* distribution of random coefficients F (parametric or non-parametric), the market share function $s_t(\delta_t; F)$ is real analytic with respect to $\delta_t \in \mathbb{R}^{C_{t1}}$.[22]

Theorem 3. *For any F, $s_t(\delta_t; F)$ is real analytic with respect to δ_t in $\mathbb{R}^{C_{t1}}$, for $t = 1, ..., T$.*

Proof. See Appendix 7.9. $\square$

While our local identification results and Lemma 2 do not hinge on the nature of the variation in $\delta_{t\mathbf{J}_t}$, the second part of our global identification argument depends on whether the variation in $\delta_{t\mathbf{J}_t}$ is *exogenous* across markets: price endogeneity restricts the variation in $\delta_{t\mathbf{J}_t}$ and leads to additional difficulties in global identification. To overcome these difficulties, we propose the use of mild restrictions on the price-setting model. In what follows, we treat separately the simpler case of exogeneous variation in $\delta_{t\mathbf{J}_t}$ and the more complex case of price endogeneity.

4.1 Exogenous Average Utilities

Here, we consider the case of exogenous variation in $\delta_{t\mathbf{J}_t}$ across markets. Given Lemma 2, denote the finite set of solutions to $m(\Sigma'; \mathbf{T}_0) = 0$ in Θ_Σ by $S(\Sigma) = \{\Sigma_r : r = 0, ..., R\}$, where Σ_0 represents the true value of Σ. On the basis of Lemma 1, define the corresponding Γ^r for $r = 0, 1, ..., R$. The real analytic property of $s_t(\delta_t'; \Sigma')$ enables to eliminate the extra solutions Σ_r, $r = 1, ..., R$, by exploiting the additional variation provided by $\delta_{t^+\mathbf{J}_{t^+}}$ for $t^+ \in \mathbf{T} \setminus \mathbf{T}_0$. Intuitively, the real analytic property guarantees that $S(\Sigma)$ is non-singleton, i.e. lack of identification, only on a union of R zero measure sets of $\delta_{t^+\mathbf{J}_{t^+}}$, $t^+ \in \mathbf{T} \setminus \mathbf{T}_0$. Because the union of any *finite* zero measure sets is still of zero measure, the real analytic property—combined with Lemma 2—ensures global identification almost everywhere given the additional variation provided by $\delta_{t^+\mathbf{J}_{t^+}}$, $t^+ \in \mathbf{T} \setminus \mathbf{T}_0$. Remember that $\delta_t(\Gamma) = (\delta_{t\mathbf{J}_t}, (\delta_{t\mathbf{b}}(\Gamma_\mathbf{b}))_{\mathbf{b} \in \mathbf{C}_{t2}})$, where $\delta_{t\mathbf{b}}(\Gamma_\mathbf{b}) = \sum_{j \in \mathbf{b}} \delta_{tj} + \Gamma_\mathbf{b}$, and define for each $r = 1, ..., R$:

$$\begin{aligned} \Delta_r^{\mathrm{ID}} = \{(\delta_{t\mathbf{J}_t})_{t \in \mathbf{T} \setminus \mathbf{T}_0}, \delta_{t\mathbf{J}_t} \in \mathbb{R}^{J_t} : \exists\, t \in \mathbf{T} \setminus \mathbf{T}_0 \text{ such that } s_t(\delta_t(\Gamma^0); \Sigma_0) \neq s_t(\delta_t'(\Gamma^r); \Sigma_r) \\ \text{for any } \delta_{t\mathbf{J}_t}' \in \mathbb{R}^{J_t}\} \end{aligned} \tag{10}$$

and $\Delta^{\mathrm{ID}} = \cap_{r=1}^R \Delta_r^{\mathrm{ID}}$. Denote by $\delta_{t\mathbf{J}_t}^0$ the true value of $\delta_{t\mathbf{J}_t}$ for which $s_t(\delta_t(\Gamma^0); \Sigma_0) = \mathscr{s}_t$.

[22] A function $f : \mathscr{X} \to \mathbb{R}$ is *real analytic* in $\mathscr{X}$ if for each $x_0 \in \mathscr{X}$, there exists a neighbourhood U of x_0 such that $f(x)$ is equal to its Taylor expansion $\sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n$ for any $x \in U$.

Theorem 4. *Suppose Assumptions 1–3 hold and Θ_Σ is compact. Then, it follows that:*

- *If $(\delta^0_{t\mathbf{J}_t})_{t\in\mathbf{T}\setminus\mathbf{T}_0} \in \Delta^{ID}$, system (9) has a unique solution in Θ_Σ and model (3) is globally identified.*
- *If $\Delta^{ID}_r \neq \emptyset$ for $r = 1, ..., R$, the Lebesgue measure of $\times_{t\in\mathbf{T}\setminus\mathbf{T}_0}\mathbb{R}^{J_t} \setminus \Delta^{ID}$ is zero.*

Proof. See Appendix 7.10. □

While the first result of Theorem 4 provides sufficient conditions for global identification, the second underlines their practical usefulness. The set Δ^{ID} is "very large" and will include the true $(\delta^0_{t\mathbf{J}_t})_{t\in\mathbf{T}\setminus\mathbf{T}_0}$ in "almost all" cases: global identification will be achieved almost everywhere.

4.2 Endogenous Prices

In what follows, we extend our global identification results from Theorem 4 to the case of endogenous prices, where the variation in $\delta_{t\mathbf{J}_t}$ across markets is restricted by the price-setting behaviour of firms. To achieve global identification, we add mild restrictions on the price-setting model. We assume the existence of exogenous cost shifters that are unobserved to the econometrician but identifiable from observed market shares and prices. Berry and Haile (2014) rely on a similar restriction (Assumption 7b, p. 1769) to obtain global identification of a simultaneous system of demand and supply by instrumental variables. However, because of the specific utility structure of model (3) under Assumption 2, our global identification argument is different and—as in the case of exogenous average utilities—does not require the use of instrumental variables.

Similar to BLP, we decompose the average utility into two additively separable parts: $\delta_{tj}(\Delta\delta_{tj}, \xi_{tj}) = \Delta\delta_{tj} + \xi_{tj}$, with $\Delta\delta_{tj}$ the portion that includes the endogenous price (and other observable exogenous characteristics) and ξ_{tj} a residual. Define the vector of such residuals in market t by $\xi_{t\mathbf{J}_t} = (\xi_{tj})_{j\in\mathbf{J}_t} \in \mathbb{R}^{J_t}$ and by $c_{t\mathbf{J}_t} = (c_{tj})_{j\in\mathbf{J}_t} \in \mathbb{R}^{J_t}_+$ a vector of cost shifters, one for each of the products in market t. An example of these cost shifters is the marginal costs of the products sold in market t. Importantly, both $\xi_{t\mathbf{J}_t}$ and $c_{t\mathbf{J}_t}$ are assumed to be unobserved to the econometrician. Endogeneity arises whenever prices—and consequently the $\Delta\delta_{tj}$'s—are chosen by firms on the basis of the residuals $\xi_{t\mathbf{J}_t}$. As for the case of exogenous average utilities, we propose an instrument-free identification strategy: a characterization of the set of $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$ that suffices for the global identification of the structural parameters $(\delta_{t\mathbf{J}_t}, \Gamma, \Sigma)$.[23]

To simplify exposition, in what follows we omit dependence on $\Delta\delta_{t\mathbf{J}_t}$. Let $D_{t\xi} \times D_{tc}$ denote the domain of $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$ for $t \in \mathbf{T}$. Suppose that the firms in market t choose prices according to pure components given the true (Γ^0, Σ_0) and $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}) \in D_{t\xi} \times D_{tc}$ (see Remark 2 below).

[23] Even though we rely on the additive separability of δ_{tj} into $\Delta\delta_{tj}$ and ξ_{tj}, the target of our identification is still δ_{tj} rather than $\Delta\delta_{tj}$ and ξ_{tj} separately.

Denote the vector of observed prices by $p_{t\mathbf{J}_t}$ and the set of equilibrium prices given $\xi_{t\mathbf{J}_t}$ and $c_{t\mathbf{J}_t}$ by $p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}) \subseteq \mathbb{R}_+^{J_t}$, given $\xi_{t\mathbf{J}_t}$ by $\mathbf{P}_t(\xi_{t\mathbf{J}_t}) = \cup_{c_{t\mathbf{J}_t} \in D_{tc}} p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$, and the grand collection of all possible equilibrium prices by $\mathbf{P}_t = \cup_{\xi_{t\mathbf{J}_t} \in D_{t\xi}} \mathbf{P}_t(\xi_{t\mathbf{J}_t})$. The vector of observed prices is an equilibrium of the price-setting model, so that $p_{t\mathbf{J}_t} \in p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$. We make the following assumption on the price-setting model.

Assumption 4.

- *D_{tc} is open in $\mathbb{R}^{J_t}$ for $t \in \mathbf{T}$.*
- *(Identifiability of Cost Shifters) $c_{t\mathbf{J}_t}$ is a C^1 function of $(\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t}) \in \{(\xi'_{t\mathbf{J}_t}, p'_{t\mathbf{J}_t}) : \xi'_{t\mathbf{J}_t} \in D_{t\xi}, p'_{t\mathbf{J}_t} \in \mathbf{P}_t(\xi_{t\mathbf{J}_t})\}$: $c_{t\mathbf{J}_t} = \phi_t(\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t})$.*

Remark 2. *Assumption 4 reflects pure components pricing: each firm chooses the prices of the individual products owned and the price of any bundle is given by the sum of the prices of its components. However, it can easily be modified to accommodate alternative pricing strategies such a mixed bundling (with a different cost shifter for each bundle): each firm chooses one price for each bundle it sells and the price of any bundle of products owned by different firms is the sum of the prices of its components. For details, see Adams and Yellen (1976).*

The second part of Assumption 4 resembles Assumption 7b by Berry and Haile (2014) and is the key to our global identification argument under prices endogeneity. Berry and Haile (2014) show that their Assumption 7b is implied by a variety of common price-setting models of oligopoly with differentiated products (see Remark 1, p. 1766). Their result follows from the assumption of "connected substitutes" on the demand system (see Definition 1, p. 1759): loosely speaking, this rules out any negative cross-price elasticity between any two products. While model (3) satisfies the connected substitutes property at the bundle-level, the property may not hold at the product-level (i.e., individual products may be complements). It then follows that, even though Remark 1 by Berry and Haile (2014) applies to the case of mixed bundling, it does not to the case of pure components pricing with complementary products.

Remark 3. *By combining the bundle-level connected substitutes property with the specific utility structure of model (3) under Assumption 2, in Appendix 7.11 we show that Assumption 4 is satisfied by several commonly employed pure components pricing models. We show that Assumption 4 is consistent with any number of firms (monopoly, duopoly, or oligopoly) playing a complete information simultaneous Bertrand-Nash game with any profile of demand synergies among products (substitutability and/or complementarity). Importantly, Assumption 4 leaves the cardinality of $p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$ unrestricted: the price-setting model is allowed to have a unique, several, or even infinitely many equilibria.*

Denote by $s_t(\delta'_t(\Gamma'); p'_{t\mathbf{J}_t}, \Sigma')$ the market share function in market t with prices $p'_{t\mathbf{J}_t} = (p'_{tj})_{j\in\mathbf{J}_t}$ and structural parameters $(\delta'_{t\mathbf{J}_t}, \Gamma', \Sigma')$. Given Lemma 2, similar to the previous section, define for each $r = 1, ..., R$:

$$\begin{aligned}\Xi^{\text{ID}}_r = \{(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})_{t\in\mathbf{T}\setminus\mathbf{T}_0} \in \times_{t\in\mathbf{T}\setminus\mathbf{T}_0}[D_{t\xi} \times D_{tc}] : &\exists\, t \in \mathbf{T} \setminus \mathbf{T}_0 \text{ such that} \\ &s_t(\delta_t(\Gamma^0); p_{t\mathbf{J}_t}, \Sigma_0) \neq s_t(\delta'_t(\Gamma^r); p_{t\mathbf{J}_t}, \Sigma_r) \\ &\text{for any } p_{t\mathbf{J}_t} \in p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}) \text{ and } \xi'_{t\mathbf{J}_t} \in D_{t\xi}\}\end{aligned}$$

and $\Xi^{\text{ID}} = \cap_{r=1}^R \Xi^{\text{ID}}_r$. In addition, define for each $r = 1, ..., R$:

$$\begin{aligned}\Xi^{\text{ID}}_{tr}(c_{t\mathbf{J}_t}) = \{\xi_{t\mathbf{J}_t} \in D_{t\xi} : &s_t(\delta_t(\Gamma^0); p_{t\mathbf{J}_t}, \Sigma_0) \neq s_t(\delta'_t(\Gamma^r); p_{t\mathbf{J}_t}, \Sigma_r) \\ &\text{for any } p_{t\mathbf{J}_t} \in p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}) \text{ and } \xi'_{t\mathbf{J}_t} \in D_{t\xi}\}.\end{aligned}$$

Denote by $(\xi^0_{t\mathbf{J}_t}, c^0_{t\mathbf{J}_t}, p^0_{t\mathbf{J}_t})$ the true value of $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}, p_{t\mathbf{J}_t})$ for which $s_t(\delta_t(\Gamma^0); p^0_{t\mathbf{J}_t}, \Sigma_0) = \mathcal{s}_t$.

Theorem 5. *Suppose Assumptions 1–4 hold and Θ_Σ is compact. Then, it follows that:*

- *If $(\xi^0_{t\mathbf{J}_t}, c^0_{t\mathbf{J}_t})_{t\in\mathbf{T}\setminus\mathbf{T}_0} \in \Xi^{ID}$, system (9) has a unique solution in Θ_Σ and model (3) is globally identified.*
- *If for any $r = 1, ..., R$, there exists $t \in \mathbf{T} \setminus \mathbf{T}_0$, for any $c_{t\mathbf{J}_t} \in D_{tc}$, such that $\Xi^{ID}_{tr}(c_{t\mathbf{J}_t}) \neq \emptyset$, then the Lebesgue measure of $\times_{t\in\mathbf{T}\setminus\mathbf{T}_0}[D_{t\xi} \times D_{tc}] \setminus \Xi^{ID}$ is zero.*

Proof. See Appendix 7.12. □

As in Theorem 4, the first result of Theorem 5 provides sufficient conditions for global identification but in the more complex case of price endogeneity. The second result, instead, stresses the practical usefulness of the proposed sufficient conditions. The set Ξ^{ID} is "very large" and will include the true $(\xi^0_{t\mathbf{J}_t}, c^0_{t\mathbf{J}_t})_{t\in\mathbf{T}\setminus\mathbf{T}_0}$ in "almost all" cases: global identification will be achieved almost everywhere.

5 Estimation

In this section, we demonstrate that our conditions for identification also guarantee desirable asymptotic properties of the Maximum Likelihood Estimator (MLE) of model (3) under Assumption 2. Several empirical papers estimate demand for bundles by MLE (e.g., Gentzkow (2007), Grzybowski and Verboven (2016), Kwak, Duvvuri, and Russell (2015)) and our result confirms the validity of this often used approach. Even though theoretically attractive, in practice the MLE of model (3) is subject to a challenge of dimensionality even under Assumption 2: the number of demand parameters can be too large to be handled (see Berry, Khwaja, Kumar,

Musalem, Wilbur, Allenby, Anand, Chintagunta, Hanemann, Jeziorski, et al. (2014)). In the spirit of BLP, we alleviate this challenge of dimensionality by proposing a convenient algorithm that breaks down the MLE's numerical search into two simpler steps that can be solved sequentially. The proposed algorithm is distinct from the classic BLP contraction mapping, and follows from a novel demand inversion result specific to Gentzkow (2007)'s model. Because of the particular utility structure of model (3) under Assumption 2, given any Γ' and Σ', the *marginal* market shares of each product $j \in \mathbf{J}_t$ (i.e., the sum of the market shares of all bundles that include $j \in \mathbf{J}_t$) are mapped to a unique $\delta'_{t\mathbf{J}_t} = (\delta'_{t1}, ..., \delta'_{tj}, ..., \delta'_{tJ_t})$. As a consequence, the proposed demand inverse effectively allows us to concentrate $(\delta'_{1\mathbf{J}_1}, ..., \delta'_{T\mathbf{J}_T})$ out of the likelihood function and to reduce the MLE's numerical search from $(\delta'_{1\mathbf{J}_1}, ..., \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma')$ to (Γ', Σ'). We refer to this estimator as to the *concentrated* MLE and show that our identification conditions ensure also its consistency and asymptotic normality.

Remark 4. *We limit our discussion to the estimation of model (3) for the case of exogenous average utilities. When Assumptions 1–4 hold and the true $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})_{t\in\mathbf{T}\setminus\mathbf{T}_0} \in \Xi^{ID}$, because of our instrument-free identification strategy, the estimation results presented below will also hold for the case of price endogeneity with no modification. In other words, also in the presence of price endogeneity, the estimation of $(\delta'_{1\mathbf{J}_1}, ..., \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma')$ will not require the use of instruments.*

5.1 Maximum Likelihood Estimation (MLE)

Denote by $I_{t\mathbf{b}}$ the number of individuals in market t observed to choose bundle $\mathbf{b}$ and by $\hat{\jmath}_{t\mathbf{b}} = \frac{I_{t\mathbf{b}}}{I}$ the corresponding observed market share. The log-likelihood function evaluated at $(\delta'_{1\mathbf{J}_1}, ..., \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma')$ can be written as:

$$\ell_I(\delta'_{1\mathbf{J}_1}, ..., \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma'; \hat{\jmath}_1, ..., \hat{\jmath}_T) = \sum_{t=1}^{T} \sum_{\mathbf{b}\in\mathbf{C}_t} \hat{\jmath}_{t\mathbf{b}} \log s_{t\mathbf{b}}(\delta'_t(\Gamma'); \Sigma'), \tag{11}$$

where $\hat{\jmath}_t = (\hat{\jmath}_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_t}$ for $t = 1, ..., T$. In addition, define

$$\ell(\delta'_{1\mathbf{J}_1}, ..., \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma') = \sum_{t=1}^{T} \sum_{\mathbf{b}\in\mathbf{C}_t} \jmath_{t\mathbf{b}} \log s_{t\mathbf{b}}(\delta'_t(\Gamma'); \Sigma'),$$

where $\jmath_{t\mathbf{b}} = s_{t\mathbf{b}}(\delta_t(\Gamma); F(\cdot; \Sigma))$. Denote the domain of parameters by $\Theta = \Theta_\delta \times \Theta_\Gamma \times \Theta_\Sigma$, where Θ_δ, Θ_Γ, and Θ_Σ are compact.

Lemma 3. *If Assumptions 1–3 hold and the true $(\delta_{t\mathbf{J}_t})_{t\in\mathbf{T}\setminus\mathbf{T}_0} \in \Delta^{ID}$, then the true $(\delta_{1\mathbf{J}_1}, ..., \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$ is the unique maximizer of $\ell(\delta'_{1\mathbf{J}_1}, ..., \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma')$ in Θ.*

Proof. See Appendix 7.13. □

Define the MLE as

$$(\hat{\delta}_{1\mathbf{J}_1}, ..., \hat{\delta}_{T\mathbf{J}_T}, \hat{\Gamma}, \hat{\Sigma}) \triangleq \underset{(\delta'_{1\mathbf{J}_1}, ..., \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma') \in \Theta}{\operatorname{argmax}} \ell_I(\delta'_{1\mathbf{J}_1}, ..., \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma'; \hat{\jmath}_1, ..., \hat{\jmath}_T). \tag{12}$$

To simplify notation, denote $(\delta'_{1\mathbf{J}_1}, ..., \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma')$ by $\theta' = (\theta'_\delta, \Gamma', \Sigma')$, the true parameters $(\delta_{1\mathbf{J}_1}, ..., \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$ by $\theta = (\theta_\delta, \Gamma, \Sigma)$, and the MLE $(\hat{\delta}_{1\mathbf{J}_1}, ..., \hat{\delta}_{T\mathbf{J}_T}, \hat{\Gamma}, \hat{\Sigma})$ by $\hat{\theta}$.

Theorem 6. *Suppose Assumptions 1–3 hold, the true* $(\delta_{t\mathbf{J}_t})_{t \in \mathbf{T} \setminus \mathbf{T}_0} \in \Delta^{ID}$, $\hat{\jmath}_{t\mathbf{b}} \xrightarrow{p} \jmath_{t\mathbf{b}}$ *for* $t = 1, ..., T$, $\mathbf{b} \in \mathbf{C}_t$, *and the standard regularity conditions detailed in Appendix 7.14 hold. Then:*

- *(Consistency)* $\hat{\theta} \xrightarrow{p} \theta$.
- *(Asymptotic Normality) There exists a matrix* $W > 0$ *such that* $\sqrt{I}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, W)$.

Proof. See Appendix 7.14. □

5.2 Concentrated MLE

Several empirical papers estimate demand for bundles by MLE (e.g., Gentzkow (2007), Grzybowski and Verboven (2016), Kwak, Duvvuri, and Russell (2015)) and our Theorems 6 confirms the validity of this often used approach whenever the full column rank condition holds. While these papers show that MLE may work well in some applications, a well known bottleneck in implementation is the possibly large number of parameters to be estimated (see Berry, Khwaja, Kumar, Musalem, Wilbur, Allenby, Anand, Chintagunta, Hanemann, Jeziorski, and Mele (2014)). As an example, suppose that in every market there are J products and individuals purchase bundles of size K. Without further restrictions, model (3) under Assumption 2 would imply J^K demand synergy parameters Γ, P parameters for the distribution of random coefficients Σ, and $J \times T$ average utility parameters $(\delta_{1\mathbf{J}_1}, ..., \delta_{T\mathbf{J}_T})$. The estimation of $J^K + P + J \times T$ parameters may be hard, especially because identification requires a large T.

A practical way to deal with this dimensionality challenge is to reduce the number of demand parameters by imposing additional ex-ante restrictions (see Kwak, Duvvuri, and Russell (2015)). To avoid arbitrary ex-ante restrictions on preferences, we reduce the dimensionality of the MLE's numerical search—rather than the number of demand parameters—by a novel demand inversion result specific to Gentzkow (2007)'s model (which differs from the classic Berry (1994)'s demand inversion) that concentrates $(\delta'_{1\mathbf{J}_1}, ..., \delta'_{T\mathbf{J}_T})$ out of the likelihood function. The proposed concentrated MLE effectively reduces the numerical search from $(\delta'_{1\mathbf{J}_1}, ..., \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma')$, i.e. $J^K + P + J \times T$ parameters, to (Γ', Σ'), i.e. $J^K + P$ parameters.

Invertibility of Marginal Market Shares. Define the observed *marginal* market share of each product $j \in \mathbf{J}_t$ as $\jmath_{tj.} = \sum_{\mathbf{b} \in \mathbf{C}_{t1}: j \in \mathbf{b}} \jmath_{t\mathbf{b}}$ and denote the collection of $\jmath_{tj.}$ across all

products in market t by $\mathcal{s}_{t\mathbf{J}_t\cdot} = (\mathcal{s}_{tj\cdot})_{j\in\mathbf{J}_t}$. Similarly, define the marginal market share function of each product $j \in \mathbf{J}_t$ as $s_{tj\cdot}(\delta'_{t\mathbf{J}_t};\Gamma',\Sigma') = \sum_{\mathbf{b}\in\mathbf{C}_{t1}:j\in\mathbf{b}} s_{t\mathbf{b}}(\delta'_t(\Gamma');\Sigma')$ and denote the collection of $s_{tj\cdot}(\delta'_{t\mathbf{J}_t};\Gamma',\Sigma')$ across all products in market t by $s_{t\mathbf{J}_t\cdot}(\cdot;\Gamma',\Sigma') = (s_{tj\cdot}(\cdot;\Gamma',\Sigma'))_{j\in\mathbf{J}_t}$. Denote the collection of demand synergies that can rationalize the observed marginal market shares $\mathcal{s}_{t\mathbf{J}_t\cdot}$ in market t by $\bar{\Theta}^t_\Gamma(\Sigma') = \{\Gamma' : \exists \delta'_{t\mathbf{J}_t} \in \mathbb{R}^{J_t} \text{ such that } s_{t\mathbf{J}_t\cdot}(\delta'_{t\mathbf{J}_t};\Gamma',\Sigma') = \mathcal{s}_{t\mathbf{J}_t\cdot}\}$ and across all T markets by $\bar{\Theta}_\Gamma = \cup_{\Sigma'\in\Theta_\Sigma} \cap_{t=1}^T \bar{\Theta}^t_\Gamma(\Sigma')$. Define also $\bar{\Theta}_\Sigma = \{\Sigma' : \cap_{t=1}^T \bar{\Theta}^t_\Gamma(\Sigma') \neq \emptyset\}$. Then, the next Assumption guarantees some minimal coherence between the observed marginal market shares and model (3) under Assumption 2.

Assumption 5. *The true $\Gamma \in \Theta_\Gamma$ and $\Theta_\Gamma \subseteq \bar{\Theta}_\Gamma$. Moreover, $\Theta_\Sigma \subseteq \bar{\Theta}_\Sigma$.*

As shown in Appendix 7.15, Assumption 2 already ensures that the true $\Gamma \in \bar{\Theta}_\Gamma$. In this sense, $\Gamma \in \Theta_\Gamma$ will always be compatible with $\Theta_\Gamma \subseteq \bar{\Theta}_\Gamma$ and Assumption 5 is well defined.

Theorem 7. *If Assumptions 1, 2, and 5 hold, then for any $\Sigma' \in \Theta_\Sigma$ and $\Gamma' \in \cap_{t=1}^T \bar{\Theta}^t_\Gamma(\Sigma') \subseteq \Theta_\Gamma$, there exists a unique $\delta'_{t\mathbf{J}_t}$ such that $s_{t\mathbf{J}_t\cdot}(\delta'_{t\mathbf{J}_t};\Gamma',\Sigma') = \mathcal{s}_{t\mathbf{J}_t\cdot}$.*

Proof. See Appendix 7.15. □

In what follows, we denote the bijection that maps $\mathcal{s}_{t\mathbf{J}_t\cdot}$ to $\delta'_{t\mathbf{J}_t}$ by $\delta'_{t\mathbf{J}_t} = \tilde{\delta}_{t\mathbf{J}_t}(\mathcal{s}_{t\mathbf{J}_t\cdot};\Gamma',\Sigma')$. Note that Theorem 7 differs from the classic Berry (1994)'s demand inversion (then generalized by Berry, Gandhi, and Haile (2013)). Berry (1994) establishes a bijection between the *full* vector of observed market shares and the *full* vector of average utilities. We rely on this classic demand inversion result throughout the paper and, for completeness, we adapt it to our context at the beginning of our analysis in Lemma 1. Differently, Theorem 7 establishes a bijection between a *transformation* of the vector of observed market shares of bundles—the *marginal* market shares of individual products—and a *sub*-vector of average utilities—the average utilities of individual products. Theorem 7 hinges on the relationship between the average utilities of bundles and of individual products specific to Gentzkow (2007)'s model, and does not immediately extend to more general demand systems. In the remainder of this section, we illustrate how Theorem 7 can be used to greatly simplify the practical implementation of the MLE of demand for bundles.

Estimator. Given Theorem 7, we define the *concentrated* MLE as

$$\begin{aligned}(\hat{\Gamma}^c,\hat{\Sigma}^c) &\triangleq \text{argmax}_{(\Gamma',\Sigma')\in\Theta_\Gamma\times\Theta_\Sigma} \ell^c_I(\Gamma',\Sigma';\hat{\mathcal{s}}_1,...,\hat{\mathcal{s}}_T)\\ &= \text{argmax}_{(\Gamma',\Sigma')\in\Theta_\Gamma\times\Theta_\Sigma} \ell_I((\tilde{\delta}_{t\mathbf{J}_t}(\hat{\mathcal{s}}_{t\mathbf{J}_t\cdot};\Gamma',\Sigma'))_{t=1,...,T},\Gamma',\Sigma';\hat{\mathcal{s}}_1,...,\hat{\mathcal{s}}_T),\\ \hat{\delta}^c_{t\mathbf{J}_t} &\triangleq \tilde{\delta}_{t\mathbf{J}_t}(\hat{\mathcal{s}}_{t\mathbf{J}_t\cdot};\hat{\Gamma}^c,\hat{\Sigma}^c),\ t=1,...,T.\end{aligned} \tag{13}$$

To simplify notation, denote $(\hat{\delta}^c_{1\mathbf{J}_1},...,\hat{\delta}^c_{T\mathbf{J}_T})$ by $\hat{\theta}^c_\delta$ and $(\hat{\theta}^c_\delta,\hat{\Gamma}^c,\hat{\Sigma}^c)$ by $\hat{\theta}^c$.

Theorem 8. *Given the same assumptions and conditions of Theorem 6 and Assumption 5:*

- *(Consistency)* $\hat{\theta}^c \xrightarrow{p} \theta$.

- *(Asymptotic Normality) There exist matrices* $W_1, W_2 > 0$ *such that* $\sqrt{I}(\hat{\theta}^c_\delta - \theta_\delta) \xrightarrow{d} \mathcal{N}(0, W_1)$ *and* $\sqrt{I}[(\hat{\Gamma}^c, \hat{\Sigma}^c) - (\Gamma, \Sigma)] \xrightarrow{d} \mathcal{N}(0, W_2)$.

Proof. See Appendix 7.16. $\square$

Implementation. In the spirit of BLP, the demand inversion result from Theorem 7 enables to break down the simultaneous numerical search for $(\delta_{1\mathbf{J}_1}, ..., \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$ implied by (12) into two simpler steps that can be solved sequentially while implementing (13):

Step 1. For any given guess of (Γ', Σ') and each market $t = 1, ..., T$, compute $\delta'_{t\mathbf{J}_t} = \tilde{\delta}_{t\mathbf{J}_t}(\hat{\jmath}_{t\mathbf{J}_t\cdot}; \Gamma', \Sigma')$ by numerically solving system $s_{t\mathbf{J}_t\cdot}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma') = \hat{\jmath}_{t\mathbf{J}_t\cdot}$. Compute the derivative $\frac{\partial \tilde{\delta}_{t\mathbf{J}_t}}{\partial(\Gamma', \Sigma')} = -\left[\frac{\partial s_{t\mathbf{J}_t\cdot}}{\partial \delta'_{t\mathbf{J}_t}}\right]^{-1}\left[\frac{\partial s_{t\mathbf{J}_t\cdot}}{\partial(\Gamma', \Sigma')}\right]$.

Step 2. Plug $\delta'_{t\mathbf{J}_t}$ for $t = 1, ..., T$ from Step 1 into $\ell_I((\delta'_{t\mathbf{J}_t})_{t=1,...,T}, \Gamma', \Sigma'; \hat{\jmath}_1, ..., \hat{\jmath}_T)$ and obtain $\ell^c_I(\Gamma', \Sigma'; \hat{\jmath}_1, ..., \hat{\jmath}_T)$. Compute its derivative with respect to (Γ', Σ'), $\frac{\partial \ell^c_I}{\partial(\Gamma', \Sigma')} = \sum_{t=1}^T \frac{\partial \ell_I}{\partial \delta'_{t\mathbf{J}_t}} \frac{\partial \tilde{\delta}_{t\mathbf{J}_t}}{\partial(\Gamma', \Sigma')} + \frac{\partial \ell_I}{\partial(\Gamma', \Sigma')}$. Check whether the current guess of (Γ', Σ') numerically maximizes $\ell^c_I(\Gamma', \Sigma'; \hat{\jmath}_1, ..., \hat{\jmath}_T)$. If yes, the current value of the parameters is $\hat{\theta}^c$. If not, use $\frac{\partial \ell^c_I}{\partial(\Gamma', \Sigma')}$ to numerically search for a new guess of (Γ', Σ') and go back to Step 1.

Finally, Theorem 8 establishes that also $\hat{\theta}^c$, similar to $\hat{\theta}$—albeit numerically less involved—, will have desirable asymptotic properties.

6 Empirical Application

We apply our methods to investigate the welfare implications of bundle-by-bundle pricing in the ready-to-eat (RTE) cereal industry in the USA. As we illustrate below, in approximately 20% of the shopping trips in our data, households are observed to purchase two or more different brands of RTE cereals. Importantly, our data record purchases rather than consumption: simultaneous purchases of different RTE cereal brands, and consequently demand synergies, may be motivated beyond complementarity in consumption. For example, Hendel (1999) and Dubé (2004) suggest that if households go shopping less often than they consume, then preference for variety can lead them to purchase multiple brands on each shopping trip in anticipation of several consumption occasions. Alternatively, if households face shopping costs for each visit to a store, one-stop shopping may be preferred to multi-stop shopping (see Thomassen, Smith, Seiler, and Schiraldi (2017)). Demand synergies in our model are catch-all parameters that can reflect complementarity in consumption, Hendel (1999)'s preference for variety, or shopping

costs.[24] Beyond the exact reasons why households purchase bundles, we note that in our data they do so and that ignoring this phenomenon altogether may lead to incorrect estimates and policy recommendations.

On the supply side of the industry, producers are observed to price their RTE cereals brand-by-brand (i.e., pure components pricing), so that the price of any bundle of brands corresponds to the sum of the prices of each brand. On the one hand, given the pronounced purchases of bundles by households, bundle-by-bundle pricing (i.e., mixed bundling pricing) can be more profitable than pure components pricing due to a better ability to price discriminate (see Armstrong (2016)). On the other hand, as noticed by Anderson and Leruth (1993) and Thanassoulis (2007), the possibility to purchase bundles of brands owned by *different* producers introduces a negative externality that can undermine the profit gains of mixed bundling.[25] Whether mixed bundling pricing would ultimately benefit or harm RTE cereal producers in the USA depends on which of the two mechanisms prevailed given households' preferences and producers' costs.

6.1 Data and Definitions

We use household-level and store-level IRI data on ready-to-eat (RTE) cereals for the period 2008-2011 for the city of Pittsfield in the USA. We report a succinct description of the data used and refer the reader to Bronnenberg, Kruger, and Mela (2008) for a more thorough discussion.

We focus on the $I = 2897$ households who are observed to purchase RTE cereals at least once from 2008 until 2011. For these households, we observe some demographics (e.g., income and family size) and a panel of shopping trips $r = 1, ..., 756665$ to 7 different grocery stores over a period of 208 weeks. A shopping trip is defined as a purchase occasion of a household to a grocery store in a given day. Each shopping trip records all the Universal Product Codes (UPCs) purchased by a household across all product categories sold by the store: during 83258 of these, RTE cereals are observed to be purchased. We define a market as a store-week combination $t = 1, ..., 1431$. Market shares are computed over the shopping trips observed in each store-week combination.

Over the sample period, the households are observed to purchase 1173 different UPCs of RTE cereals. In order to make the empirical analysis feasible, we reduce the number of different RTE cereal products by collecting UPCs into what we call brands. We define $J = 16$ different brands on the basis of producers and ingredients. We classify producers into six groups: General Mills, Kellogg's, Quaker, Post, Small Producers, and Private Labels. The first four correspond to the four largest RTE cereal producers, "Small Producers" correspond to the

[24] In particular, our model can rationalize shopping costs with positive demand synergies. In addition, as we illustrate in Appendix 7.1, the model of preference for variety by Hendel (1999) and Dubé (2004) is a special case of ours with negative demand synergies.

[25] See footnote 4 for details

remaining producers, and "Private Labels" correspond to the UPCs directly branded by the retailers (i.e., the stores). We collect the UPCs of each of the producers into three categories on the basis of their ingredients: fiber and/or whole grains, added sugar, and regular. Table 1 lists these RTE cereal brands and their average market shares across the shopping trips with *some* RTE cereal purchase. We use the store-level data to compute brand-level prices for each brand j and store-week combination t, p_{tj}. Each p_{tj} is computed as the average price per 16oz across the UPCs belonging to brand j in store-week t.

Table 1: RTE Cereal Brands and Market Shares

RTE Cereal Brands		Average Market Shares
General Mills	Fiber/Whole Grain	34.99%
Kellogg's	Regular	8.46%
	Fiber/Whole Grain	17.30%
	Added Sugar	4.45%
Quaker	Regular	1.42%
	Fiber/Whole Grain	9.09%
	Added Sugar	0.76%
Post	Regular	0.04%
	Fiber/Whole Grain	8.31%
	Added Sugar	0.69%
Private Labels	Regular	3.21%
	Fiber/Whole Grain	3.12%
	Added Sugar	2.01%
Small Producers	Regular	0.14%
	Fiber/Whole Grain	4.30%
	Added Sugar	1.71%

Notes: The Table lists the 16 RTE cereal brands obtained by aggregating UPCs as described in the text. For each brand, we report its average market share across the 83258 shopping trips with *some* RTE cereal purchases.

During each shopping trip r, a household i is considered to purchase RTE cereal brand j whenever they are observed to purchase at least a UPC of brand j. Households are considered to purchase multi-product bundles only when purchasing at least two *different* brands of RTE cereals during the *same* shopping trip. In our view, this is a conservative measure of households'

demand for bundles.[26]

As Figure 1 illustrates, in more than 10% of the shopping trips with *some* RTE cereal purchase, single-person households are observed to purchase two or more *different* brands of RTE cereals. For computational convenience, we focus our analysis on the shopping trips with observed purchases of at most two different RTE cereal brands, thus discarding 3.28% of the shopping trips with some RTE cereal purchase. In 17.69% of the 83258 shopping trips with some RTE cereal purchase, households are observed to purchase two different brands of RTE cereals. Figure 1 also shows that multi-person households have a higher demand for bundles than single-person households. Table 2 further expands on this point: as family size increases, the average bundle size purchased by households steadily increases.

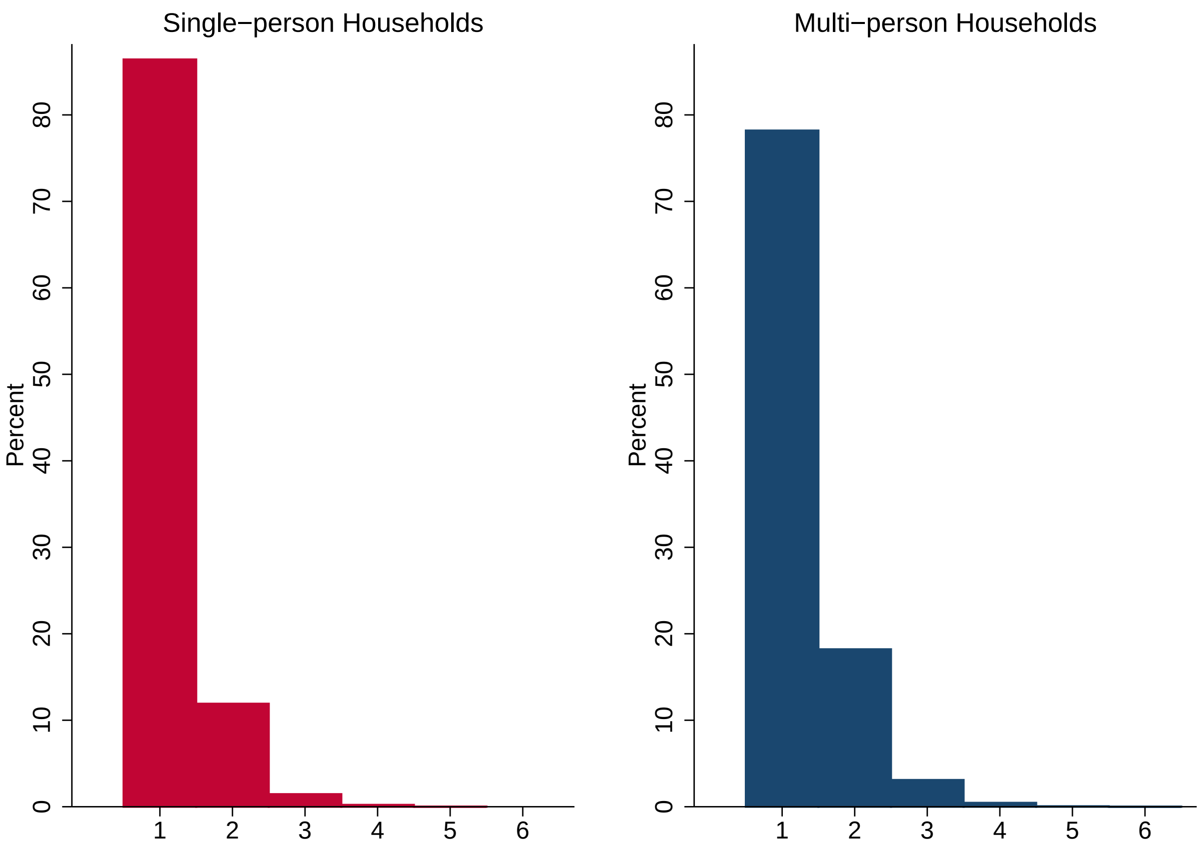
We make the standard assumption that RTE cereal purchases do not determine store choice and take store choice as given in our econometric model. We consider household i to choose the outside option $j = 0$ in store-week combination t whenever no RTE cereal brand is purchased during shopping trip r in t (in general, *something* must be purchased for a shopping trip to be in the data). Around 89% of all shopping trips do not involve any purchase of RTE cereals.

We construct choice sets at the level of store-week combination t: any household during any shopping trip taking place in t is assumed to face choice set $\mathbf{C}_t$. Choice set $\mathbf{C}_t$ is made of three components: single-product bundles, multi-product bundles of size 2, and the outside option. From the store-level data, we observe which of the 16 brands of RTE cereals are available in each store-week combination t. Denote this set of available brands in t by $\mathbf{J}_t$. In addition, households can purchase multi-product bundles $(j_1, j_2) \in (\mathbf{J}_t \times \mathbf{J}_t)\backslash\{(k_1, k_2)|k_1 = k_2\}$: bundles made of pairs of *different* RTE cereal brands. Finally, during any shopping trip, households may decide not to purchase any RTE cereal brand, which we denote by $j = 0$. By combining these purchase possibilities, the choice set faced during all shopping trips taking place in t is $\mathbf{C}_t = \mathbf{J}_t \cup \{0\} \cup (\mathbf{J}_t \times \mathbf{J}_t)\backslash\{(k_1, k_2)|k_1 = k_2\}$.[27]

[26]For instance, the purchases of different RTE cereal brands across different shopping trips during the same week are considered as independent purchases of individual brands rather than multi-product bundles. To avoid confounding factors due to quantity discounts, we do not count as bundles the purchases of multiple units of the same brand within the same shopping trip. Accommodating either less conservative definitions of bundles or purchases of multiple units of the same brand would not represent any conceptual challenge for the proposed methods.

[27]The choice set $\mathbf{C}_t$ also excludes those multi-product bundles that are not chosen during any shopping trip in t. Even though all brands in $\mathbf{J}_t$ have positive market shares by construction, some combination of brands (j_1, j_2) from $(\mathbf{J}_t \times \mathbf{J}_t)\backslash\{(k_1, k_2)|k_1 = k_2\}$ may not be observed to be jointly purchased.

Figure 1: # *of RTE Cereal Brands* per Household in Shopping Trip



Notes: US IRI household-level data. Each observation is a shopping trip for a household, where we use purchase decisions in the RTE cereal category for the period 2008-2011.

Table 2: Descriptive Statistics, by Family Size

Family Size	#Households	Ave. Bundle Size
1	732	1.12
2	1184	1.16
3	423	1.20
4	359	1.22
5	139	1.25
6	52	1.26
≥ 7	8	1.37

Notes: The Table shows the distribution of "family size" among the $I = 2897$ households in our data. For each level of family size, we report the average number of different RTE cereal brands observed to be purchased per shopping trip by the corresponding households.

6.2 Model Specification

Any household i is observed going on several shopping trips, each taking place in a specific store-week combination t (our definition of market). The indirect utility of household i from shopping trip r in market t by choosing single-product bundle $j \in \mathbf{J}_t$ is:

$$
\begin{aligned}
U_{irtj} &= u_{itj} + \varepsilon_{irtj} \\
&= \delta_{tj} + \mu_{itj} + \varepsilon_{irtj},
\end{aligned}
\tag{14}
$$

$$
\mu_{itj} = -p_{tj} \exp(d_i \alpha + v_i) + \eta_{ij}
$$

where $u_{itj} = \delta_{tj} + \mu_{itj}$, δ_{tj} is market t-specific average utility for RTE cereal brand $j \in \mathbf{J}_t$, μ_{itj} is a household i-specific utility deviation from δ_{tj}, and ε_{irtj} is an idiosyncratic error term. p_{tj} is the price of brand j in store-week combination t, and $d_i\alpha + v_i$ is a vector of household i-specific price coefficients made of two components: an observable part that is a function of the household demographics d_i and an unobserved random component v_i. η_{ij} is an unobserved household i-specific preference for brand j, which is constant across i's shopping trips and potentially correlated across brands.

Specification (14) encapsulates the entire effect of price p_{tj} in the household i-specific μ_{itj}. In terms of the notation used in section 4.2, this implies that $\Delta\delta_{tj} = 0$ and that $\delta_{tj} = \xi_{tj}$. Even though we use household-level data, we will face price endogeneity whenever the producer of RTE cereal brand j sets price p_{tj} by taking the average utility δ_{tj} into consideration. Our proposed estimators essentially address this endogeneity problem by treating the average utility δ_{tj} for each brand j in each market t as a *fixed effect*: the standard MLE will directly estimate all of these fixed effects along with the other structural parameters, while the concentrated MLE will concentrate them out from the numerical search, greatly simplifying implementation.

The indirect utility of household n from shopping trip r in market t by choosing multi-product bundle $\mathbf{b}$ is:

$$
\begin{aligned}
U_{irt\mathbf{b}} &= \sum_{j \in \mathbf{b}} u_{itj} + \Gamma_{i\mathbf{b}} + \varepsilon_{irt\mathbf{b}} \\
&= \sum_{j \in \mathbf{b}} (\delta_{tj} + \mu_{itj}) + \Gamma_{\mathbf{b}} + \zeta_{i\mathbf{b}} + \varepsilon_{irt\mathbf{b}} \\
&= \sum_{j \in \mathbf{b}} \delta_{tj} + \Gamma_{\mathbf{b}} + \left[\sum_{j \in \mathbf{b}} \mu_{itj} + (d_i\gamma + \tilde{\zeta}_{i\mathbf{b}}) \right] + \varepsilon_{irt\mathbf{b}} \\
&= \delta_{t\mathbf{b}} + \mu_{it\mathbf{b}} + \varepsilon_{irt\mathbf{b}},
\end{aligned}
\tag{15}
$$

where $\delta_{t\mathbf{b}} = \sum_{j \in \mathbf{b}} \delta_{tj} + \Gamma_{\mathbf{b}}$ is market t-specific average utility for bundle $\mathbf{b}$, $\mu_{it\mathbf{b}}$ is household i-specific utility deviation from $\delta_{t\mathbf{b}}$, $\Gamma_{i\mathbf{b}}$ is household i-specific demand synergy among the

brands in bundle $\mathbf{b}$, and $\varepsilon_{irt\mathbf{b}}$ is an idiosyncratic error term. The demand synergy parameter $\Gamma_{i\mathbf{b}} = \Gamma_{\mathbf{b}} + \zeta_{i\mathbf{b}}$ captures the extra utility household i obtains from buying the RTE cereal brands in bundle $\mathbf{b}$ jointly rather than separately. It is the sum of $\Gamma_{\mathbf{b}}$, common to all households, and $\zeta_{i\mathbf{b}} = d_i\gamma + \tilde{\zeta}_{i\mathbf{b}}$, where $d_i\gamma$ is a function of observed household demographics d_i and $\tilde{\zeta}_{i\mathbf{b}}$ is an unobserved random component. Finally, the indirect utility of household i from shopping trip r in market t by choosing the outside option is assumed to be:

$$U_{irt0} = \varepsilon_{irt0}. \tag{16}$$

Suppose that ε_{irt0} and the $\varepsilon_{irt\mathbf{b}}$'s are i.i.d. Gumbel. Express $\mu_{it\mathbf{b}} = \mu_{it\mathbf{b}}(\beta_i)$ as a function of the unobservable $\beta_i = (v_i, \eta_i, \tilde{\zeta}_i) = (v_i, (\eta_{ij})_{j\in\mathbf{J}}, (\tilde{\zeta}_{i\mathbf{b}})_{\mathbf{b}\in\mathbf{B}})$.[28] Then, given β_i and $(\delta_{t\mathbf{J}_t}, \alpha, \gamma, \Gamma) = ((\delta_{tj})_{j\in\mathbf{J}_t}, \alpha, \gamma, (\Gamma_{\mathbf{b}})_{\mathbf{b}\in\mathbf{B}})$, household i's choice probability of bundle $\mathbf{b} \in \mathbf{C}_t$ during shopping trip r in market t is:

$$s_{irt\mathbf{b}}(\delta_{t\mathbf{J}_t}, \alpha, \gamma, \Gamma; \beta_i) = \frac{e^{\delta_{t\mathbf{b}} + \mu_{it\mathbf{b}}(\beta_i)}}{\sum_{\mathbf{b}'\in\mathbf{C}_t} e^{\delta_{t\mathbf{b}'} + \mu_{it\mathbf{b}'}(\beta_i)}}.$$

We assume $\beta_i = (v_i, \eta_i, \tilde{\zeta}_i)$ to be normally distributed and denote its c.d.f. by $\Phi(\cdot; \Sigma)$. Let $y_{itr\mathbf{b}} \in \{0, 1\}$ be an indicator for whether household i purchased bundle $\mathbf{b}$ during shopping trip r in market t, with $\sum_{\mathbf{b}\in\mathbf{C}_t} y_{itr\mathbf{b}} = 1$. Let T_i denote the set of markets for which we observe shopping trips by household i. For each $t \in T_i$, define R_{it} as the set of shopping trips by household i that took place in market t. By integrating over the distribution of β_i, we obtain the likelihood of i's observed choices $y_i = (y_{itr\mathbf{b}})_{t\in T_i, r\in R_{it}, \mathbf{b}\in\mathbf{C}_t}$:

$$L_i(\delta_{1\mathbf{J}_1}, ...\delta_{T\mathbf{J}_T}, \alpha, \gamma, \Gamma, \Sigma; y_i) = \int \prod_{t\in T_i} \prod_{r\in R_{it}} \prod_{\mathbf{b}\in\mathbf{C}_t} (s_{irt\mathbf{b}}(\delta_{t\mathbf{J}_t}, \alpha, \gamma, \Gamma; \beta_i))^{y_{itr\mathbf{b}}} \, d\Phi(\beta_i; \Sigma).$$

By aggregating over the $I = 2897$ households, the likelihood function for the entire set of observed choices is:

$$L_I(\delta_{1\mathbf{J}_1}, ...\delta_{T\mathbf{J}_T}, \alpha, \gamma, \Gamma, \Sigma; y_1, ..., y_I) = \prod_{i=1}^{2897} L_i(\delta_{1\mathbf{J}_1}, ...\delta_{T\mathbf{J}_T}, \alpha, \gamma, \Gamma, \Sigma, y_i). \tag{17}$$

Finally, we estimate the demand parameters $(\delta_{1\mathbf{J}_1}, ...\delta_{T\mathbf{J}_T}, \alpha, \gamma, \Gamma, \Sigma)$ on the basis of the concentrated MLE derived from likelihood function (17).[29] To get a sense of the practical advantages implied by Theorem 7, in the current application the concentrated MLE reduces the numerical search from approximately 20000 to the more manageable 133 parameters $(\alpha, \gamma, \Gamma, \Sigma)$.

[28] $\mathbf{J}$ and $\mathbf{B}$ are defined as, respectively, the union of all $\mathbf{J}_t$ and of all $(\mathbf{J}_t \times \mathbf{J}_t) \backslash \{(k_1, k_2) | k_1 = k_2\}$ for $t = 1, ..., T$.

[29] Even though (17) is expressed in terms of individual choices $(y_1, ..., y_I)$ rather than of sampled market shares $(\hat{\jmath}_1, ..., \hat{\jmath}_T)$, it can be easily shown that it satisfies the conditions of Theorem 8.

6.3 Estimates of Demand for RTE cereal Bundles

We estimate the price coefficients that characterize the distribution of the price sensitivity $-\exp(d_i\alpha + v_i)$: the vector of income-specific price coefficients α and the standard deviation σ_v of the random coefficient v_i, distributed i.i.d. normal. d_i is a vector of 12 mutually exclusive dummies indicating the level of income earned by household i (as coded in the IRI data), so that $d_i\alpha$ accounts for the possibility that wealthier households may be less sensitive to prices.[30] The estimated price coefficients are presented in Table 3. The average price sensitivity $\mathbb{E}[-\exp(d_i\alpha + v_i)]$ across all households is estimated to be -0.558 and its standard deviation to 0.581, suggesting the presence of unobserved heterogeneity in price sensitivity across households. Wealthier households do not seem to be systematically less price sensitive, but this can be motivated by their potentially larger family sizes.

Table 3: Estimated Price Coefficients

	Price Coefficient by Income Level												SD of v_i
	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	α_{10}	α_{11}	α_{12}	σ_v
Coef.	−0.66	−0.62	−0.73	−0.57	−0.63	−0.55	−0.64	−0.63	−0.60	−0.70	−0.61	−0.65	0.28
Std Err.	0.010	0.013	0.013	0.012	0.010	0.008	0.007	0.0010	0.008	0.012	0.007	0.009	0.003

Notes: The Table reports the estimated price coefficients that characterize the distribution of the price sensitivity $-\exp(d_i\alpha + v_i)$: the vector of income-specific price coefficients α and the standard deviation σ_v of the random coefficient v_i, which is distributed i.i.d. normal. The income level of each household belongs to one of 12 classes. The classes are numbered according to the increasing level of income.

We specify the RTE cereal brand-specific random intercepts as $\eta_{ij} = \eta_{i,\text{type}} + \tilde{\eta}_{ij}$, where $\eta_{i,\text{type}}$ captures household i's unobserved and correlated preferences across cereal types $\{R, F/W, S\}$ and $\tilde{\eta}_{ij}$ captures i.i.d. unobserved preference for brand j.[31] $\eta_{i,\text{type}}$ and $\tilde{\eta}_{ij}$ are assumed to be normally distributed and mutually independent. Following Gentzkow (2007), we also include $\eta_{i,\text{type}} \in \{\eta_{i,R}, \eta_{i,F/W}, \eta_{i,S}\}$ to the random intercept of each brand j (on the basis of its ingredients) so that any two brands with similar ingredients will share the same $\eta_{i,\text{type}}$, and allow $\{\eta_{i,R}, \eta_{i,F/W}, \eta_{i,S}\}$ to be freely correlated.

Table 4 reports the estimated parameters of the joint normal distribution of η_{ij}. The estimates underline the importance of controlling for this form of unobserved heterogeneity: all the estimates of the standard deviation and correlation parameters are statistically significant. In particular, households' unobserved preferences for healthier F/W and children S cereal brands are positively correlated, while unobserved preferences for regular R cereal brands seem to correlate negatively with both F/W and S cereals brands.

[30] The 12 income classes are numbered according to an increasing level of income.

[31] Remember that cereal type R refers to "Regular," F/W to "Fiber/Whole Grain," and S to "Added Sugar."

Table 4: Estimated Distribution of Random Intercepts, η_{ij}

	SD of $\eta_{i,\text{type}}$			Correlation of $\eta_{i,\text{type}}$			SD of $\tilde{\eta}_{ij}$
	σ_{η_R}	$\sigma_{\eta_{F/W}}$	σ_{η_S}	$\text{corr}_{R,F/W}$	$\text{corr}_{R,S}$	$\text{corr}_{F/W,S}$	$\sigma_{\tilde{\eta}}$
Coef.	0.09	0.52	0.96	-0.84	-0.81	0.74	0.90
Std. Err.	0.014	0.007	0.016	0.008	0.012	0.013	0.005

Notes: The Table reports the estimated distribution of the random intercepts $\eta_{ij} = \eta_{i,\text{type}} + \tilde{\eta}_{ij}$, where $\eta_{i,\text{type}}$ captures household i's unobserved and correlated preferences across cereal types $\{R, F/W, S\}$ and $\tilde{\eta}_{ij}$ captures i.i.d. residual unobserved preference for brand j. $\eta_{i,\text{type}}$ and $\tilde{\eta}_{ij}$ are mutually independent and distributed normal. Cereal type R refers to "Regular," F/W to "Fiber/Whole Grain," and S to "Added Sugar."

We specify the demand synergies associated to bundle $\mathbf{b}$ as $\Gamma_{i\mathbf{b}} = \Gamma_{\mathbf{b}} + \tilde{\zeta}_{i\mathbf{b}}$ for single-person households and as $\Gamma_{i\mathbf{b}} = \Gamma_{\mathbf{b}} + \gamma_k + \tilde{\zeta}_{i\mathbf{b}}$ for multi-person households of family size k, with $k = 2, ..., 7$. The parameter γ_k captures systematic differences between the average demand synergies of households with family size k and single-person households (whose average demand synergies are simply $\Gamma_{\mathbf{b}}$). Differently, $\tilde{\zeta}_{i\mathbf{b}}$ represents a i-specific i.i.d. normal unobserved deviation from $\Gamma_{\mathbf{b}} + \gamma_k$. This specification is meant to rationalize the purchase patterns documented in Table 2: larger families have higher demand levels for bundles of different brands of RTE cereals. The estimated demand synergies are reported in Tables 5 (for the $\Gamma_{\mathbf{b}}$'s) and 6 (for the γ_k's).

As Table 5 illustrates, there is a substantial amount of positive demand synergies among different brands of RTE cereals ($\Gamma_{\mathbf{b}} > 0$).[32] Importantly for competition policy, as we will see in the next section, there appear to be positive demand synergies not only among brands within producer, but also among brands between producers. For example, the first column shows that single-person households exhibit significant positive demand synergies between General Mills and 11 out of the 15 remaining brands. Moreover, in line with the evidence from Table 2, the demand synergy shifters γ_k's are all estimated to be positive and increasing in family size, k. As a household's family size increases, the strength of the positive demand synergies among different RTE cereal brands increases. Intuitively, larger families may be more likely to purchase different brands in order to satisfy increasingly heterogeneous RTE cereal tastes within the household (e.g., adults and children of different ages). The standard deviation $\sigma_{\tilde{\zeta}}$ of the random coefficient $\tilde{\zeta}_{n\mathbf{b}}$ is estimated to be small but significant, suggesting the presence of household-specific heterogeneity in demand synergies beyond differences in family size.

[32] See the Table notes for an interpretation of the missing values.

Table 5: Estimated Average Demand Synergy Parameters, $\Gamma_{\mathbf{b}}$

		G.Mills	Kellogg's			Quaker			Post			Private			Small Producers		
		F/W	R	F/W	S	R	F/W	S	R	F/W	S	R	F/W	S	R	F/W	S
G. Mills	F/W	·															
Kellogg's	R	0.88*	·														
	F/W	0.83*	1.78*	·													
	S	0.87*	2.08*	2.12*	·												
Quaker	R	0.79	1.89*	−1.13	1.25	·											
	F/W	0.85*	1.05*	0.89*	0.73	0.21	·										
	S	0.86*	0.96*	0.96*	1.30*	2.39*	2.54*	·									
Post	R	1.48	2.23*	1.74	·	·	3.51*	·	·								
	F/W	0.82*	0.71*	0.80*	0.61*	1.43	1.19*	0.89*	3.82*	·							
	S	0.77*	0.29	1.00*	0.49	·	0.67	1.25	·	2.02*	·						
Private	R	0.81*	0.54*	0.48*	−0.08	1.10	1.88*	0.41	·	0.92*	0.52	·					
	F/W	0.36*	0.31	0.32*	0.08	·	1.69*	−0.12	3.33*	0.72*	0.82	2.52*	·				
	S	0.44*	0.49	0.67*	−0.08	0.87	0.77*	1.83	·	0.84*	1.59*	3.01*	2.96*	·			
Small P.	R	0.13	−0.04	0.42*	0.28	1.40	0.32	0.70	·	0.05	·	0.66	0.10	−0.07	·		
	F/W	0.87*	0.38*	0.34*	0.07	·	0.31*	0.25	·	0.43*	−0.05	0.41*	0.84*	0.94*	2.58*	·	
	S	0.15	0.17	0.44	−0.25	·	1.32	0.15	·	0.58	1.04	1.43	1.20	1.76*	1.80*	2.83*	·

Notes: The Table reports the estimated average demand synergies for all bundles of RTE cereal brands for single-person households. Cereal type R refers to "Regular," F/W to "Fiber/Whole Grain," and S to "Added Sugar." An off-diagonal missing value "·" refers to an observed zero market share for the corresponding bundle (we do not observe the purchase of the bundle in any of the shopping trips in our data). Differently, missing values "·" along the diagonal refer to our definition of bundle: we only consider as bundles the joint purchases of *different* brands of RTE cereals. The $*$ refers to the estimate being significantly different from zero at a level $\geq 95\%$.

Table 6: Estimated Demand Synergy Shifters

	Demand Synergy by Family Size						Std. Dev.
	γ_2	γ_3	γ_4	γ_5	γ_6	$\gamma_{\geq 7}$	$\sigma_{\tilde{\zeta}}$
Coef.	0.25	0.33	0.40	0.55	0.54	0.50	0.04
Std Err.	0.017	0.024	0.023	0.033	0.056	0.117	0.010

Notes: The Table reports the estimated demand synergy shifters parameters embedded in $\zeta_{i\mathbf{b}} = d_i\gamma + \tilde{\zeta}_{i\mathbf{b}}$. More precisely, a separate γ_k for all those households with family size $k > 1$ and the standard deviation $\sigma_{\tilde{\zeta}}$ of the i.i.d. normally distributed unobserved component $\tilde{\zeta}_{i\mathbf{b}}$. The average demand synergies of single-person households from Table 5 are the reference for these shifters.

Table 7 shows the average (across markets) estimated own- and cross-price elasticities of demand. Each entry reports the percent change in the *marginal* market share of the column RTE cereal brand with respect to a 1% increase in the price of the row RTE cereal brand. Given the estimated market share function $\hat{s}_{t\mathbf{b}}$ for each bundle $\mathbf{b} \in \mathbf{C}_{t1}$ in market t, the estimated marginal market share function of brand $j \in \mathbf{J}_t$ is defined as $\hat{s}_{tj.} = \sum_{\mathbf{b}\in\mathbf{C}_{t1}:j\in\mathbf{b}} \hat{s}_{t\mathbf{b}}$. We consider two brands to be complements (substitutes) whenever their compensated cross-price elasticity of demand is negative (positive).[33] Standard models of demand for individual brands rule out the possibility of complementarity among different brands and restrict cross-price elasticities to be positive. However, Table 7 suggests that, for the RTE cereal industry in the US, this restriction may be too stringent: complementarity among different brands appear to be pervasive. For example, the first column shows that households exhibit various degrees of complementarity between General Mills and 12 out of the 15 remaining brands. According to intuition, complementarity is more pronounced among those brands with larger positive demand synergies ($\Gamma_{\mathbf{b}} > 0$). In the next section, we investigate the economic importance of accounting for complementarity in demand estimation by estimating restricted versions of the demand model and by comparing the corresponding counterfactual results.

[33] On the various definitions of complementarity and substitutability see Samuelson (1974) and, more recently, Manzini, Mariotti, and Ülkü (2018).

Table 7: Average Estimated Own- and Cross-Price Elasticities

		G.Mills	Kellogg's			Quaker			Post			Private			Small Producers		
		F/W	R	F/W	S	R	F/W	S	R	F/W	S	R	F/W	S	R	F/W	S
G.Mills	F/W	-1.745	-0.014	-0.040	-0.006	-0.001	-0.009	-0.003	-0.000	-0.026	-0.002	-0.006	-0.000	0.000	0.001	-0.034	0.001
Kellogg's	R	-0.066	-1.997	-0.133	-0.037	-0.003	-0.008	-0.002	-0.001	-0.010	0.001	-0.000	0.002	0.001	0.003	0.003	0.001
	F/W	-0.078	-0.055	-1.629	-0.051	0.001	-0.007	-0.003	-0.001	-0.019	-0.003	-0.000	0.001	-0.000	0.000	0.002	0.001
	S	-0.057	-0.067	-0.221	-1.694	-0.001	-0.000	-0.008	0.001	-0.002	0.002	0.006	0.006	0.008	0.002	0.019	0.003
Quaker	R	-0.061	-0.068	0.031	-0.012	-2.952	0.003	-0.032	0.001	-0.052	0.005	-0.009	0.007	-0.002	-0.009	0.029	0.004
	F/W	-0.069	-0.013	-0.029	-0.000	0.000	-1.616	-0.034	-0.006	-0.040	0.000	-0.004	-0.027	-0.000	0.001	0.007	-0.003
	S	-0.068	-0.009	-0.033	-0.019	-0.005	-0.092	-1.714	0.001	-0.018	-0.007	0.002	0.007	-0.001	-0.000	0.012	0.002
Post	R	-0.043	-0.037	-0.039	0.015	0.001	-0.111	0.005	-1.721	-0.410	0.005	0.008	-0.068	0.006	0.008	0.042	0.002
	F/W	-0.082	-0.006	-0.031	-0.000	-0.002	-0.016	-0.002	-0.010	-1.650	-0.020	-0.006	-0.004	-0.002	0.002	-0.004	0.000
	S	-0.048	0.007	-0.034	0.006	0.001	0.001	-0.007	0.001	-0.125	-1.951	0.002	-0.002	-0.020	0.006	0.022	-0.001
Private	R	-0.057	-0.001	-0.002	0.006	-0.001	-0.005	0.001	0.001	-0.019	0.000	-1.353	-0.064	-0.067	-0.001	0.001	-0.003
	F/W	-0.004	0.004	0.007	0.007	0.001	-0.032	0.003	-0.005	-0.011	-0.001	-0.062	-1.204	-0.078	0.002	-0.021	-0.002
	S	0.002	0.003	-0.005	0.012	-0.000	-0.001	-0.001	0.001	-0.012	-0.012	-0.097	-0.116	-1.177	0.004	-0.019	-0.008
Small P.	R	0.015	0.009	0.000	0.004	-0.002	0.002	-0.001	0.001	0.008	0.004	-0.002	0.003	0.004	-2.325	-0.283	-0.005
	F/W	-0.093	0.001	0.001	0.006	0.001	0.002	0.002	0.001	-0.004	0.003	0.000	-0.006	-0.004	-0.052	-1.922	-0.024
	S	0.041	0.010	0.015	0.018	0.001	-0.012	0.005	0.001	0.004	-0.002	-0.010	-0.006	-0.024	-0.013	-0.359	-1.568

Notes: The Table reports the average estimated own- (diagonal) and cross-price (off-diagonal) elasticities, where averages are computed across markets. Each entry reports the percent change in the *marginal* market share of the column RTE cereal brand with respect to a 1% increase in the price of the row RTE cereal brand. Given the estimated market share function $\hat{s}_{t\mathbf{b}}$ for each bundle $\mathbf{b} \in \mathbf{C}_{t1}$ in market t, the estimated marginal market share function of any brand $j \in \mathbf{J}_t$ is defined as $\hat{s}_{tj.} = \sum_{\mathbf{b}\in\mathbf{C}_{t1}:j\in\mathbf{b}} \hat{s}_{t\mathbf{b}}$. Cereal type R refers to "Regular," F/W to "Fiber/Whole Grain," and S to "Added Sugar."

6.4 Counterfactuals on Pricing Strategies and Market Structures

The ability of households to purchase bundles of brands may allow producers to improve price discrimination and to raise profits through non-linear pricing (see Armstrong (2016)). However, as noticed by Anderson and Leruth (1993) and Thanassoulis (2007) in the case of duopoly, the possibility of purchasing bundles of brands owned by *different* producers introduces a negative externality that may undermine the profit gains due to non-linear pricing.[34] Whether non-linear pricing in the form of mixed bundling would ultimately benefit or harm the RTE cereal industry in the USA depends on which of the two mechanisms prevailed given households' preferences and producers' costs. Building on the demand estimates from the previous section, in what follows we explore the welfare implications of mixed bundling pricing and their interactions with market structure.

6.4.1 Bundling Externality and Pricing Strategy as Commitment Device

To provide some guidance for our counterfactual simulations, in the current section we investigate how the magnitude of the negative "bundling externality" relates both to market structure and to the producers' pricing strategy (pure components and mixed bundling pricing).

Denote a producer by f, the set of brands f owns by $\mathbf{J}_f$, and the collection of five producers except f by $-f$. Pure components restricts the price of any bundle (j, j') with $j, j' \in \mathbf{J}_f$ to be the sum of the prices of the individual brands j and j', so that $p_{(j,j')} = p_j + p_{j'}$. In the case of monopoly, producer f owns all the RTE cereal brands in the market, which therefore belong to $\mathbf{J}_f$, and chooses all the prices $p_f = (p_j, p_{(j,j')})_{j,j' \in \mathbf{J}_f, j' \neq j}$. In this scenario, the profit obtained by f from choosing prices p_f is:

$$\begin{aligned}\pi_f(p_f) &= \sum_{j \in \mathbf{J}_f} s_j(p_f)(p_j - c_j) + \sum_{j,j' \in \mathbf{J}_f, j \neq j'} s_{(j,j')}(p_f)(p_{(j,j')} - c_j - c_{j'}) \\ &= \pi_f^1(p_f) + \pi_f^2(p_f),\end{aligned} \tag{18}$$

where $s_j(p_f)$ is the market share of the single-product bundle j (i.e., made of those shopping trips where j is the only brand purchased), while $s_{(j,j')}(p_f)$ is the market share of the multi-product bundle (j, j') with $j, j' \in \mathbf{J}_f$, $j \neq j'$. c_j denotes the marginal cost of brand j and $c_j + c_{j'}$ the marginal cost of bundle (j, j'). More succinctly, $\pi_f^1(p_f)$ refers to the profit from sales of single-product bundles and $\pi_f^2(p_f)$ refers to the profit from sales of multi-product bundles. Because $\mathbf{J}_f$ includes all the RTE cereal brands available in the market, the monopolist will fully internalize the ramifications of bundling when choosing p_f. Differently, when f does not own all the brands in the market, households' ability to purchase multi-product bundles made of brands owned by *different* producers gives rise to a negative externality that can limit the

[34] See footnote 4.

profitability of the industry.

In the case of oligopoly (including duopoly), given any prices $p_{-f} = \{(p_j, p_{(j,j')})_{j,j' \in \mathbf{J}_{f'}, j' \neq j} : f' \in -f\}$ by the other producers, the profit obtained by producer f from choosing prices $p_f = (p_j, p_{(j,j')})_{j,j' \in \mathbf{J}_f, j' \neq j}$ is:

$$
\begin{aligned}
\pi_f(p_f, p_{-f}) &= \sum_{j \in \mathbf{J}_f} s_j(p_f, p_{-f})(p_j - c_j) + \sum_{j,j' \in \mathbf{J}_f, j \neq j'} s_{(j,j')}(p_f, p_{-f})(p_{(j,j')} - c_j - c_{j'}) \\
&\quad + \sum_{j \in \mathbf{J}_f} s_{j,-f}(p_f, p_{-f})(p_j - c_j) \\
&= \pi_f^1(p_f, p_{-f}) + \pi_f^2(p_f, p_{-f}) + \pi_f^3(p_f, p_{-f}),
\end{aligned}
\tag{19}
$$

where $s_j(p_f, p_{-f})$ is the market share of the single-product bundle j (i.e., made of those shopping trips where j is the only brand purchased), $s_{(j,j')}(p_f, p_{-f})$ is the market share of the multi-product bundle (j, j'), $j \neq j'$, with both j and j' owned by f, and

$$
s_{j,-f}(p_f, p_{-f}) = \sum_{k \in \mathbf{J}_{f'}, f' \in -f} s_{(j,k)}(p_f, p_{-f}) \tag{20}
$$

is the sum of the market shares of the multi-product bundles made of j and each of the other brands owned by $-f$. The main difference between (18) and (19) is the additional term $\pi_f^3(p_f, p_{-f})$, the profit from sales of multi-product bundles made of a brand owned by f and the other owned by $f' \in -f$, which gives rise to what we call the bundling externality.

When a competing producer $f' \in -f$ increases the price of individual brand $k' \in \mathbf{J}_{f'}$—for given p_f—both $\pi_f^1(p_f, p_{-f})$ and $\pi_f^2(p_f, p_{-f})$ will increase. This follows from the market shares $s_j(p_f, p_{-f})$ and $s_{(j,j')}(p_f, p_{-f})$ being increasing in $p_{k'}$. Differently, an increase in $p_{k'}$ by $f' \in -f$ will give rise to two opposing forces acting on $\pi_f^3(p_f, p_{-f})$, which may result in $\pi_f^3(p_f, p_{-f})$ to decrease. On the one hand, it will induce a reduction in the denominator of $s_{j,-f}(p_f, p_{-f})$: the denominator of each market share $s_{(j,k)}(p_f, p_{-f})$ from (20) will decrease. On the other, however, it will also lead to a decrease in the numerator of $s_{(j,k')}(p_f, p_{-f})$ given that bundle (j, k'), with price $p_{(j,k')} = p_j + p_{k'}$, becomes more expensive. Overall, an increase in $p_{k'}$ may then bring to a net reduction in $s_{(j,k')}(p_f, p_{-f})$ and, consequently, in $s_{j,-f}(p_f, p_{-f})$. Note that, similar to any other bundle $(j, k) \in \{\mathbf{J}_f \times \mathbf{J}_{f'} : f' \in -f\}$, $s_{(j,k')}(p_f, p_{-f})$ enters both the demand faced by f and that faced by f'. Even though f' does not fully take this into account when choosing prices, any decision about $p_{k'}$ will directly affect also f through the market shares of all bundles (j, k') with $j \in \mathbf{J}_f$. This is what we refer to as bundling externality.

The strength of the bundling externality relates both to market structure and to the producers' pricing strategy. The relationship with market structure can be readily seen from (19) and (20): the smaller the set of brands owned by f relative to the total set of brands in the market, the larger the number of market shares of multi-product bundles included in $\pi_f^3(p_f, p_{-f})$ rather

than in $\pi_f^2(p_f, p_{-f})$.[35] The extreme cases being $\pi_f^3(p_f, p_{-f}) = 0$ when f is a monopolist and $\pi_f^2(p_f, p_{-f}) = 0$ when f only owns one brand. With $\pi_f^3(p_f, p_{-f}) = 0$, there will be no room for the bundling externality, while with $\pi_f^2(p_f, p_{-f}) = 0$ its extent will be maximal. The next Proposition instead sheds some light on the relationship between the bundling externality and the producers' pricing strategy.[36]

Proposition 1. *Suppose f only owns two brands, $\mathbf{J}_f = \{j, j'\}$, and denote by $p_f^* = (p_j^*, p_{j'}^*, p_{(j,j')}^* = p_j^* + p_{j'}^*)$ the prices chosen by f under pure components. Under mixed bundling, f will have an incentive to choose $p_j > p_j^*$, $p_{j'} > p_{j'}^*$, and $p_{(j,j')} < p_{(j,j')}^*$ as long as:*

$$\mathscr{E}_j^{(j,j')}\lambda_j + \mathscr{E}_{j'}^{(j,j')}\lambda_{j'} + \mathscr{E}_{(j,j')}^{(j,j')}\lambda_{(j,j')} < -1, \tag{21}$$

where $\mathscr{E}_r^{(j,j')}$ is the elasticity of $s_{(j,j')}$ with respect to p_r evaluated at p_f^* and $\lambda_r = (p_r^* - c_r)/p_r^*$ is r's Lerner index, for $r \in \{j, j', (j,j')\}$.

Proof. See Appendix 7.17. $\square$

When condition (21) is satisfied, f will have an incentive to choose $p_j > p_j^*$ so that any $p_{(j,k')} = p_j + p_{k'}$ with $j \in \mathbf{J}_f$ and $k' \in \mathbf{J}_{f'}, f' \in -f$, will be higher under mixed bundling relative to pure components. In other words, when condition (21) holds, mixed bundling will exacerbate the negative bundling externality relative to pure components pricing. In particular, condition (21) will be satisfied whenever the (negative) own-price elasticity of the bundle, $\mathscr{E}_{(j,j')}^{(j,j')}$, is "large" relative to the (positive) cross-price elasticities with respect to p_j and $p_{j'}$, $\mathscr{E}_j^{(j,j')}$ and $\mathscr{E}_{j'}^{(j,j')}$. Using our demand and marginal costs estimates, we can verify whether condition (21) holds in practice.[37] Table 8 summarizes our estimates of the left-hand side of condition (21) across markets. Condition (21) holds given our estimates of demand and marginal costs: in the counterfactuals, we should expect a stronger negative bundling externality under mixed bundling relative to pure components pricing.

[35] In the counterfactuals in which we simulate alternative market structures, we keep the total set of multi-product bundles unchanged. Essentially, we focus on the short run and assume that bundling choices (on the side of producers) are exogenous to market structure. As a consequence, for any market structure, the market share of each multi-product bundle must be included either in $\pi_f^2(p_f, p_{-f})$ or in $\pi_f^3(p_f, p_{-f})$.

[36] Our result is an extension of Armstrong (2013)'s Proposition 1 (page 455), about a monopolist selling two brands, to the case of an oligopolist selling two brands and facing mixed logit demand. Our result can easily be extended to oligopolies with producers owning more than two brands.

[37] Following Nevo (2000) and Nevo (2001), we estimate marginal costs (given our demand estimates) by assuming that the observed prices are a Nash equilibrium of a pricing game in which the six RTE cereal producers directly choose prices according to pure components pricing, abstracting from modelling the relationships between producers and retailer. The only role played by the retailer is to choose the prices of the private labels they own, similar to any of the other five competing producers.

Table 8: Median of Condition (21) across Markets

$\mathscr{E}^{(j,j')}_{(j,j')}\lambda_{(j,j')}$	+	$\mathscr{E}^{(j,j')}_{j}\lambda_{j}$	+	$\mathscr{E}^{(j,j')}_{j'}\lambda_{j'}$	=	Total
-1.725		0.013		0.012		-1.700

Notes: Within each market, we compute the average of the estimated $\mathscr{E}^{(j,j')}_{(j,j')}\lambda_{(j,j')}$, $\mathscr{E}^{(j,j')}_{j}\lambda_{j}$, and $\mathscr{E}^{(j,j')}_{j}\lambda_{j}$ across bundles, and then report the median of these averages across markets. The rightmost column reports the sum of these terms as in the left-hand side of condition (21).

6.4.2 Counterfactual Simulations

We take the observed scenario of pure components pricing and oligopolistic competition among RTE cereal producers as a reference, and simulate the changes in welfare induced by mixed bundling pricing and different market structures.[38] On the one hand, we consider four alternative market structures: "competition," where we suppose that each individual brand is owned and sold by a different (fictional) producer (for a total of 16 producers), "oligopoly," which corresponds to the observed (or factual) oligopolistic competition among the six producers, "duopoly," where we suppose that five of the producers (General Mills, Kellog's, Quaker, Post, and the Small Producers) perfectly collude and compete as one against the private labels (whose prices are chosen by the retailer), and "monopoly," where we suppose that the six producers perfectly collude or merge into a monopolist. On the other hand, we consider two pricing strategies: "pure components," where each producer chooses the prices of the individual brands they own and "mixed bundling," where each producer chooses a specific price for each individual brand and for each two-product bundle they own. Note that, in the counterfactual scenario we call "competition," there is no difference between pure components and mixed bundling pricing.

The results of these counterfactual simulations are reported in Table 9. The Table reports relative changes in prices, profits, and consumer surplus associated with each of the different counterfactual scenarios. The top and central panels of the Table report, respectively, price and profit changes associated to different types of bundles as in equations (19) and (20): $\pi^1_f(p_f, p_{-f})$ is the profit from sales of individual brands $j \in \mathbf{J}_f$, $\pi^2_f(p_f, p_{-f})$ is the profit from sales of bundles (j, j') with $j, j' \in \mathbf{J}_f, j' \neq j$ (i.e., bundles of brands owned by f), and $\pi^3_f(p_f, p_{-f})$ is the profit from sales of bundles (j, k') with $j \in \mathbf{J}_f, k' \in \mathbf{J}_{f'}, f' \in -f$ (i.e., bundles of brands owned by different producers). All percentage changes are relative to the observed scenario of oligopoly with pure components pricing. As a consequence, in displaying results across counterfactuals,

[38] Given our estimates of demand and marginal costs, we simulate each profile of counterfactual prices using the necessary first order conditions for a Nash equilibrium of the corresponding pricing game. For example, in a monopoly with mixed bundling pricing, the same agent chooses a specific price for each individual brand and for each two-product bundle so to maximize industry profits.

the definitions of $\mathbf{J}_f$ and $\mathbf{J}_{f'}$ we use are those from the observed oligopoly.[39] For example, the displayed $\pi_f^3(p_f, p_{-f})$ for any counterfactual always refers to the change in profit corresponding to the specific bundles of brands owned by different producers in the observed oligopoly.

Table 9: Counterfactual Simulations from the Full Model

Pricing Strategy		pure components pricing			mixed bundling pricing		
Market Structure	competition	oligopoly	duopoly	monopoly	oligopoly	duopoly	monopoly
Price change							
$p_j : j \in \mathbf{J}_f$	+5.47%	0%	-2.16%	-2.79%	+5.94%	+1.70%	+13.32%
$p_{(j,j')} : j, j' \in \mathbf{J}_f, j' \neq j$	+5.61%	0%	-1.90%	-2.50%	-21.08%	-16.85%	-14.03%
$p_{(j,k')} : j \in \mathbf{J}_f, k' \in \mathbf{J}_{f'}, f' \in -f$	+5.05%	0%	-2.16%	-2.75%	+5.52%	-1.56%	-14.93%
Profit change	-0.48%	0%	+0.24%	+0.26%	+0.29%	+0.73%	+4.67%
$\pi_f^1(p_f, p_{-f})$	+1.04%	0%	-0.96%	-1.10%	-0.07%	-0.88%	-3.47%
$\pi_f^2(p_f, p_{-f})$	-1.06%	0%	+0.34%	+0.48%	+21.38%	+17.05%	+14.71%
$\pi_f^3(p_f, p_{-f})$	-4.46%	0%	+3.48%	+3.89%	-4.58%	+0.59%	+23.94%
Consumer Surplus change	-5.33%	0%	+5.87%	+6.61%	-1.18%	+5.10%	+0.77%
family size= 1	-5.08%	0%	+5.80%	+6.58%	-2.01%	+4.50%	-2.65%
family size≥ 2	-5.37%	0%	+5.88%	+6.61%	-1.09%	+5.17%	+1.17%

Notes: The Table reports average counterfactual changes in prices (top panel), profits (central panel), and consumer surplus (bottom panel) when RTE cereal producers adopt different pricing strategies under alternative market structures (with respect to the observed oligopoly). Each column refers to a specific pricing-market structure combination. The second column refers to the observed oligopoly in the data. See the text for a description of the counterfactual market structures corresponding to the other columns. The top and central panels report, respectively, price and profit changes associated to different types of bundles as in equations (19) and (20): $\pi_f^1(p_f, p_{-f})$ is the profit from sales of individual brands $j \in \mathbf{J}_f$, $\pi_f^2(p_f, p_{-f})$ is the profit from sales of bundles (j, j') with $j, j' \in \mathbf{J}_f, j' \neq j$ (i.e., bundles of brands owned by f), and $\pi_f^3(p_f, p_{-f})$ is the profit from sales of bundles (j, k') with $j \in \mathbf{J}_f, k' \in \mathbf{J}_{f'}, f' \in -f$ (i.e., bundles of brands owned by different producers). All percentage changes are relative to the observed scenario of oligopoly with pure components pricing. As a consequence, in displaying results across counterfactuals, the definitions of $\mathbf{J}_f$ and $\mathbf{J}_{f'}$ we use are those from the observed oligopoly. For example, the displayed $\pi_f^3(p_f, p_{-f})$ for any counterfactual always refers to the change in profit corresponding to the specific bundles of brands owned by different producers in the observed oligopoly. The bottom panel reports consumer surplus changes for households of different family sizes.

The results for pure components pricing confirm the classic result by Cournot (1838): mergers between producers selling complementary brands can be socially desirable. In pure components, the prices of all individual brands—and consequently of all bundles—decrease as the level of competition decreases: while industry-level profit remains basically unchanged, consumer surplus increases with market concentration. As market structure becomes more concentrated, producers internalize more of the bundling externality when choosing prices, as can be seen from the relative increase in $\pi_f^3(p_f, p_{-f})$ from -4.46% for competition to $+3.89\%$ for monopoly. With mixed bundling pricing, in line with the classic result by Adams and Yellen (1976), a monopolist would maximize the industry-level profit due to the enhanced ability to price discriminate and to the absence of the negative bundling externality. Similar to the case of

[39] To be clear, we do not keep $\mathbf{J}_f$ and $\mathbf{J}_{f'}$ fixed in the simulation of different counterfactuals, but only in the way results are displayed in Table 9 (to enhance comparability).

pure components—as expected—, also with mixed bundling when market structure becomes more concentrated producers internalize more of the bundling externality: $\pi^3_f(p_f, p_{-f})$ increases steadily from -4.46% for competition to $+23.94\%$ for monopoly.

Table 9 highlights that the profit gains of mixed bundling with respect to pure components are sharply decreasing in the level of competition: $(4.67 - 0.26)\% = +4.41\%$ for monopoly, 0.49% for duopoly, and 0.29% for oligopoly (for competition this is 0% by construction). In line with the theoretical predictions from Proposition 1 (whose conditions are verified in Table 8), for any given market structure, the negative bundling externality appears to be exacerbated by mixed bundling relative to pure components: $(0.59 - 3.48)\% = -2.89\%$ for duopoly and -4.58% for oligopoly.[40] Interestingly, in the observed oligopoly, the bundling externality would play such a prominent role with mixed bundling that the industry-level profit would basically *not* be higher than in the observed case of pure components pricing. In other words, if the observed oligopoly were to opt to mixed bundling as opposed to the observed pure components pricing, the relative profit losses due to an exacerbated bundling externality would actually be as large as the relative profit gains due to an enhanced ability to price discriminate.

These results are in line with the theoretical predictions by Anderson and Leruth (1993), Thanassoulis (2007), and Armstrong (2016), which show that pure components may indeed be preferred by a duopoly to mixed bundling because of the possibly extreme pro-competitive effects of mixed bundling (what we encapsulate in the negative bundling externality). Importantly, our results ignore any cost difference in the practical implementation of the alternative pricing strategies: pure components pricing could very well be strictly preferred due to the potentially higher costs associated to the implementation of the logistically more involved mixed bundling pricing (see Chu, Leslie, and Sorensen (2011)). In addition, Table 9 shows that mixed bundling pricing would also harm consumers because of the higher levels of price discrimination: given any market structure, pure components pricing leads to larger levels of consumer surplus than mixed bundling pricing.

[40] Remember that in monopoly, by construction, there will not be any negative bundling externality irrespective of the pricing strategy.

References

Adams, W. J., and J. L. Yellen (1976): "Commodity bundling and the burden of monopoly," *The quarterly journal of economics*, pp. 475–498.

Allen, R., and J. Rehbeck (forthcoming): "Identification with additively separable heterogeneity," *Econometrica*.

Anderson, S. P., and L. Leruth (1993): "Why firms may prefer not to price discriminate via mixed bundling," *International Journal of Industrial Organization*, 11(1), 49–61.

Armstrong, M. (2013): "A more general theory of commodity bundling," *Journal of Economic Theory*, 148(2), 448–472.

——— (2016): "Nonlinear pricing," *Annual Review of Economics*, 8, 583–614.

Augereau, A., S. Greenstein, and M. Rysman (2006): "Coordination versus differentiation in a standards war: 56k modems," *The RAND Journal of Economics*, 37(4), 887–909.

Berry, S., A. Gandhi, and P. Haile (2013): "Connected substitutes and invertibility of demand," *Econometrica*, 81(5), 2087–2111.

Berry, S., A. Khwaja, V. Kumar, A. Musalem, K. Wilbur, G. Allenby, B. Anand, P. Chintagunta, M. Hanemann, P. Jeziorski, and A. Mele (2014): "Structural models of complementary choices," *Marketing Letters*, 25(3), 245–256.

Berry, S., A. Khwaja, V. Kumar, A. Musalem, K. C. Wilbur, G. Allenby, B. Anand, P. Chintagunta, W. M. Hanemann, P. Jeziorski, et al. (2014): "Structural models of complementary choices," *Marketing Letters*, 25(3), 245–256.

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

Berry, S. T., and P. A. Haile (2014): "Identification in differentiated products markets using market level data," *Econometrica*, 82(5), 1749–1797.

Bronnenberg, B. J., M. W. Kruger, and C. F. Mela (2008): "Database paper—The IRI marketing data set," *Marketing science*, 27(4), 745–748.

Chu, C. S., P. Leslie, and A. Sorensen (2011): "Bundle-size pricing as an approximation to mixed bundling," *The American Economic Review*, 101(1), 263–303.

Ciarlet, P. G. (2013): *Linear and nonlinear functional analysis with applications*, vol. 130. Siam.

Compiani, G. (2017): "Nonparametric Demand Estimation in Differentiated Products Markets," .

Cournot, A. (1838): *Researches into the Mathematical Principles of the Theory of Wealth.* Macmillan Co.

Crawford, G. S., R. S. Lee, M. D. Whinston, and A. Yurukoglu (2018): "The welfare effects of vertical integration in multichannel television markets," *Econometrica*, 86(3), 891–954.

Crawford, G. S., and A. Yurukoglu (2012): "The welfare effects of bundling in multichannel television markets," *The American Economic Review*, 102(2), 643–685.

Debreu, G. (1959): *Theory of value: An axiomatic analysis of economic equilibrium*, no. 17. Yale University Press.

Dubé, J.-P. (2004): "Multiple discreteness and product differentiation: Demand for carbonated soft drinks," *Marketing Science*, 23(1), 66–81.

Ershov, D., W. P. Laliberté, and S. Orr (2018): "Mergers in a Model with Complementarity," *Working Paper*, pp. 1–50.

Fox, J. T., and N. Lazzati (2017): "A note on identification of discrete choice models for bundles and binary games," *Quantitative Economics*, 8(3), 1021–1036.

Gale, D., and H. Nikaido (1965): "The Jacobian matrix and global univalence of mappings," *Mathematische Annalen*, 159(2), 81–93.

Gentzkow, M. (2007): "Valuing new goods in a model with complementarity: Online newspapers," *The American Economic Review*, 97(3), 713–744.

Grzybowski, L., and F. Verboven (2016): "Substitution between fixed-line and mobile access: the role of complementarities," *Journal of Regulatory Economics*, 49(2), 113–151.

Hendel, I. (1999): "Estimating multiple-discrete choice models: An application to computerization returns," *The Review of Economic Studies*, 66(2), 423–446.

Komunjer, I. (2012): "Global identification in nonlinear models with moment restrictions," *Econometric Theory*, 28(4), 719–729.

Krantz, S. G., and H. R. Parks (2002): *A primer of real analytic functions.* Springer Science & Business Media.

Kwak, K., S. D. Duvvuri, and G. J. Russell (2015): "An analysis of assortment choice in grocery retailing," *Journal of Retailing*, 91(1), 19–33.

Lewbel, A. (forthcoming): "The Identification Zoo-Meanings of Identification in Econometrics," *Journal of Economic Literature*, 7.

Lewis, A. D. (2009): "Semicontinuity of rank and nullity and some consequences," *Author' s notes on semicontinuity of rank and nullity.*

Manzini, P., M. Mariotti, and L. Ülkü (2018): "Stochastic complementarity," *The Economic Journal.*

Mas-Colell, A., M. D. Whinston, J. R. Green, et al. (1995): *Microeconomic theory*, vol. 1. Oxford university press New York.

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

Mityagin, B. (2015): "The zero set of a real analytic function," *arXiv preprint arXiv:1512.07276.*

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

——— (2001): "Measuring market power in the ready-to-eat cereal industry," *Econometrica*, 69(2), 307–342.

Newey, W. K., and D. McFadden (1994): "Large sample estimation and hypothesis testing," *Handbook of econometrics*, 4, 2111–2245.

Rothenberg, T. J. (1971): "Identification in parametric models," *Econometrica: Journal of the Econometric Society*, pp. 577–591.

Ruiz, F. J., S. Athey, and D. M. Blei (2017): "SHOPPER: A Probabilistic Model of Consumer Choice with Substitutes and Complements," *working paper.*

Samuelson, P. A. (1974): "Complementarity: An essay on the 40th anniversary of the Hicks-Allen revolution in demand theory," *Journal of Economic literature*, 12(4), 1255–1289.

Sargan, J. D. (1983): "Identification and lack of identification," *Econometrica: Journal of the Econometric Society*, pp. 1605–1633.

Sher, I., and K. i. Kim (2014): "Identifying combinatorial valuations from aggregate demand," *Journal of Economic Theory*, 153, 428–458.

Thanassoulis, J. (2007): "Competitive mixed bundling and consumer surplus," *Journal of Economics & Management Strategy*, 16(2), 437–467.

Thomassen, Ø., H. Smith, S. Seiler, and P. Schiraldi (2017): "Multi-category competition and market power: a model of supermarket pricing," *American Economic Review*, 107(8), 2308–51.

Varian, H. R. (1992): "Microeconomic analysis," *3rd. ed. ed. New York, London: WW Norton & Company.*

7 Appendix

7.1 Hendel (1999) and Dubé (2004) as Special Cases of Model (3)

In this Appendix, we illustrate that the model of multiple discreteness originally proposed by Hendel (1999) in the context of demand for computers and then applied by Dubé (2004) to the context of demand for soft drinks is a special case of model (3). In particular, Hendel (1999)'s model is a nested version of ours in which each demand synergy parameter $\Gamma_{it\mathbf{b}}$ is restricted in a special way. Hendel (1999)'s model is about individuals who go shopping less often than they consume. During any purchase occasion, individuals may buy several units of different products in anticipation of the various consumption occasions they will face before the next shopping trip. Suppose there are J different products and denote by $\mathbf{J}$ their collection. Denote by 0 the outside option, the decision of consuming none of the J products. Denote by $R_i \in \mathbb{N}$ the maximal number of units of any product that individual i can consume during any consumption occasion, and by K_i the number of consumption occasions in between any two shopping trips. On any consumption occasion, Hendel (1999) assumes that different products are perfect substitutes, so that each individual will effectively choose a certain number of units of at most *one* product j. As a consequence, the actual choice set faced by individual i on any consumption occasion can be defined as:

$$\mathbf{A}_i = \{\underbrace{(j, ..., j)}_{q} : \text{ for } j \in \mathbf{J},\ q = 1, ..., R_i\} \cup \{0\},$$

where q is the number of units of any product j that could be consumed on this consumption occasion and 0 is the outside option (the choice of not consuming anything on this consumption occasion). Then, individual i's choice set during any purchase occasion is:

$$\mathbf{C}_i = \underbrace{\mathbf{A}_i \times ... \times \mathbf{A}_i}_{K_i},$$

where each element of $\mathbf{C}_i$ is a bundle of size up to $R_i \times K_i$. To ease exposition, we represent each bundle $\mathbf{b} \in \mathbf{C}_i$ by $\mathbf{b} = (j_k, q_k)_{k=1}^{K_i}$, where (j_k, q_k) refers to the chosen product and corresponding number of units on consumption occasion k. Denote by $(j_k, q_k) = (0, 0)$ the decision of not consuming anything on consumption occasion k.

For the rest of this Appendix, we focus on Dubé (2004)'s notation, which specializes Hendel (1999)'s model to the case of demand for bundles in grocery shopping. Following Dubé (2004)'s equation (2) at page 68, denote by $(\Psi_{ij_k k} q_k)^\alpha S_i$ the indirect utility of individual i from choosing (j_k, q_k) on consumption occasion k: $\Psi_{ij_k k}$ is i's perceived quality for product j_k on consumption occasion k, S_i is an i-specific scaling factor, and $\alpha \in (0, 1)$ captures the curvature of the utility

function.[41] Moreover, denote by p_{j_k} the price of one unit of product j_k and by y_i the income of individual i. Then, from Dubé (2004)'s equation (6) at page 69, the indirect utility of individual i from choosing bundle $\mathbf{b} = ((j_1, q_1), ..., (j_{K_i}, q_{K_i})) \in \mathbf{C}_i$ is:

$$
\begin{aligned}
U_{i\mathbf{b}} &= \sum_{k=1}^{K_i} (\Psi_{ij_k k} q_k)^\alpha S_i - \sum_{k=1}^{K_i} p_{j_k} q_k + y_i \\
&= \sum_{k=1}^{K_i} (\Psi_{ij_k k} q_k)^\alpha S_i + \sum_{k=1}^{K_i} (\Psi_{ij_k k})^\alpha S_i q_k - \sum_{k=1}^{K_i} (\Psi_{ij_k k})^\alpha S_i q_k - \sum_{k=1}^{K_i} p_{j_k} q_k + y_i \\
&= \sum_{k=1}^{K_i} (\Psi_{ij_k k} q_k)^\alpha S_i + \sum_{k=1}^{K_i} \sum_{q=1}^{q_k} (\Psi_{ij_k k})^\alpha S_i - \sum_{k=1}^{K_i} (\Psi_{ij_k k})^\alpha S_i q_k - \sum_{k=1}^{K_i} \sum_{q=1}^{q_k} p_{j_k} + y_i \qquad (22) \\
&= \sum_{k=1}^{K_i} \sum_{q=1}^{q_k} [(\Psi_{ij_k k})^\alpha S_i - p_{j_k}] + \sum_{k=1}^{K_i} (\Psi_{ijk})^\alpha S_i [q_k^\alpha - q_k] + y_i \\
&= \sum_{k=1}^{K_i} \sum_{q=1}^{q_k} u_{ij_k k} + \Gamma_{i\mathbf{b}} + y_i,
\end{aligned}
$$

where $u_{ij_k k} = (\Psi_{ij_k k})^\alpha S_i - p_{j_k}$ and $\Gamma_{i\mathbf{b}} = \sum_{k=1}^{K_i} (\Psi_{ij_k k})^\alpha S_i [q_k^\alpha - q_k]$. The sum over q_k on the right hand side of (22) is zero when $q_k = 0$. Note that Dubé (2004) assumes $\Psi_{ij_k k} \geq 0$. As a consequence, the demand synergy $\Gamma_{i\mathbf{b}}$ will be constrained to be strictly negative as long as $\Psi_{ij_k k} > 0$. Dubé (2004)'s demand model is therefore a special case of model (3) with non-positive demand synergies and without the idiosyncratic error terms distributed i.i.d. Gumbel.

7.2 Some Intuition about Identification

In this Appendix, we illustrate—at an intuitive level—that even simple versions of model (3) raise non-trivial identification issues. First, we illustrate that without further restrictions on Γ_t, model (3) can hardly be identified. Second, we discuss three examples that highlight Gentzkow (2007)'s insight: when $\Gamma_t = \Gamma$, the availability of purchase data for different markets helps identification.

Suppose there are only two products in each market t, $\mathbf{J}_t = \{1, 2\}$. The indirect utility of

[41] Note that Dubé (2004)'s equation (2) at page 68 reports the *direct* utility function defined over the entire vector $(q_{jk})_{j=1}^J$ of possible units for each product $j \in \mathbf{J}$ on consumption occasion k. However, because of the assumption of perfect substitutes mentioned earlier, positive units $q_{jk} > 0$ will be chosen for at most one product j on any consumption occasion k. For this reason, here we simplify the discussion and immediately consider the *indirect* utility of choosing (j_k, q_k) with $q_{j_k k} = q_k$.

individuals in market t by choosing bundle $\mathbf{b} \in \{0, 1, 2, (1, 2)\}$ is:[42]

$$\begin{aligned}
U_{it0} &= \varepsilon_{it0}, \\
U_{it1} &= \delta_{t1} + \mu_{i1} + \varepsilon_{it1}, \\
U_{it2} &= \delta_{t2} + \mu_{i2} + \varepsilon_{it2}, \\
U_{it(1,2)} &= \delta_{t1} + \delta_{t2} + \mu_{i1} + \mu_{i2} + \Gamma_t + \varepsilon_{it(1,2)},
\end{aligned} \tag{23}$$

where $\mu_i = (\mu_{i1}, \mu_{i2})$ is distributed according to $F(\mu_i; \Sigma)$, $\Sigma = (\sigma, r)$, and $\varepsilon_{it\mathbf{b}}$ is i.i.d. Gumbel. Suppose that the econometrician observes without error the market shares $\mathscr{s}_{t\mathbf{b}}$ of each bundle $\mathbf{b} \in \{0, 1, 2, (1, 2)\}$ for each market $t = 1, ..., T$. For any given true market shares, $\mathscr{s}_t = (\mathscr{s}_{t1}, \mathscr{s}_{t2}, \mathscr{s}_{t(1,2)})$, we consider the model to be identified when the true structural parameters $(\delta_{t1}, \delta_{t2}, \Gamma_t)$ and (σ, r) represent the unique solution to the following system:

$$\begin{aligned}
&s_t(\delta'_{t1}, \delta'_{t2}, \delta'_{t(1,2)}(\Gamma'_t); \sigma', r') = \mathscr{s}_t \\
&\text{subject to } \delta'_{t(1,2)}(\Gamma'_t) - \delta'_{t1} - \delta'_{t2} = \Gamma'_t
\end{aligned} \tag{24}$$

for $t = 1, ..., T$. Note that, because of the constraint $\delta'_{t(1,2)}(\Gamma'_t) = \delta'_{t1} + \delta'_{t2} + \Gamma'_t$, knowledge of $(\delta'_{t1}, \delta'_{t2}, \Gamma'_t)$ is enough to pin down the t-specific average utility of multi-product bundle $(1, 2)$, $\delta'_{t(1,2)}(\Gamma'_t)$. Even in this simple example, a formal discussion of identification on the basis of system (24) would require to deal with cumbersome details, and these would prevent us from detecting the main mechanism at work. We then investigate the behaviour of a *linearized* version of system (24) around the true $((\delta_t(\Gamma_t))_{t=1}^T, \sigma, r)$. In the main text, we then show how the intuition from the linearized system extends to the general version of the model.

We linearize system (24) around the true $((\delta_t(\Gamma_t))_{t=1}^T, \sigma, r)$:[43]

$$\begin{aligned}
&\delta'_t(\Gamma'_t) = \delta_t(\Gamma_t) + \left.\frac{\partial s_t^{-1}}{\partial(\sigma', r')}\right|_{(\sigma', r') = (\sigma, r)} (\sigma' - \sigma, r' - r)^{\mathrm{T}} \\
&\text{subject to } \delta'_{t(1,2)}(\Gamma'_t) - \delta'_{t1} - \delta'_{t2} = \Gamma'_t
\end{aligned} \tag{25}$$

for $t = 1, ..., T$, where we denote transposition by T. Define $M = (-1, -1, 1)$ and $M\delta'_t(\Gamma'_t) = \delta'_{t(1,2)}(\Gamma'_t) - \delta'_{t1} - \delta'_{t2}$. Then, by multiplying the first line of (25) by M and by plugging in the constraint, one obtains:

$$\Gamma'_t = \Gamma_t + M \left.\frac{\partial s_t^{-1}}{\partial(\sigma', r')}\right|_{(\sigma', r') = (\sigma, r)} (\sigma' - \sigma, r' - r)^{\mathrm{T}} \tag{26}$$

[42] Bundle $\mathbf{b} = 0$ corresponds to the choice of not purchasing any product, the outside option.

[43] The basis for this linearization follows from Lemma 1. Lemma 1 shows that the *inverse* market share $s_t^{-1}(\cdot; \sigma', r')$ is a well defined function: for any given $\mathscr{s}_t$ and (σ', r') in a neighbourhood of (σ, r), there exists a unique δ'_t such that $s_t(\delta'_t; \sigma', r') = \mathscr{s}_t$. In addition, the dependence of $\delta'_t = s_t^{-1}(\mathscr{s}_t; \sigma', r')$ on (σ', r') is continuously differentiable.

for $t = 1, ..., T$. System (26) has T equations in $T+2$ unknowns, Γ'_t for $t = 1, ..., T$ and (σ', r'). The system is under-determined and (25) does not have a unique solution. One way to reduce the dimensionality of (26) is to add restrictions on Γ_t. In this paper, building on Gentzkow (2007)'s insight, we consider the case of $\Gamma_t = \Gamma$ for $t = 1, ..., T$:

$$\Gamma' = \Gamma + M \frac{\partial s_t^{-1}}{\partial(\sigma', r')}\bigg|_{(\sigma', r')=(\sigma, r)} (\sigma' - \sigma, r' - r)^{\mathrm{T}}. \tag{27}$$

Note that because Γ' and Γ in (27) are no longer market specific, system (27) has T equations in only three unknowns, Γ' and (σ', r'). By taking market 1 as a reference, one can then difference out Γ' and Γ, and the admissible (σ', r') candidates are characterized by the following linear system:

$$M \left[\frac{\partial s_t^{-1}}{\partial(\sigma', r')} - \frac{\partial s_1^{-1}}{\partial(\sigma', r')}\right]\bigg|_{(\sigma', r')=(\sigma, r)} (\sigma' - \sigma, r' - r)^{\mathrm{T}} = 0, \tag{28}$$

of $t = 2, ..., T$. If (σ', r') is a solution to system (28), then given (σ', r') one can determine the corresponding Γ' from system (27). In turn, given (σ', r') and Γ', one can obtain the remaining parameters $(\delta'_{t1}, \delta'_{t2})_{t=1}^T$ from system (25). Collectively, these $(\delta'_t(\Gamma'))_{t=1}^T$ and (σ', r') constitute a solution to system (25).

Example 1. Suppose there are two markets $T = 2$, $t \in \{a, b\}$, and that r is known to equal zero. The true structural parameters are $(\delta_{a1}, \delta_{a2}, \delta_{b1}, \delta_{b2}, \Gamma, \sigma)$ and the observed market shares are $\mathscr{s}_a = (\mathscr{s}_{a1}, \mathscr{s}_{a2}, \mathscr{s}_{a(1,2)})$ and $\mathscr{s}_b = (\mathscr{s}_{b1}, \mathscr{s}_{b2}, \mathscr{s}_{b(1,2)})$. Because r is assumed to be known and to equal zero, system (28) simplifies to the equation:

$$M \left[\frac{\partial s_b^{-1}}{\partial \sigma'} - \frac{\partial s_a^{-1}}{\partial \sigma'}\right]\bigg|_{(\sigma', r')=(\sigma, 0)} (\sigma' - \sigma) = 0, \tag{29}$$

where $M = (-1, -1, 1)$. Note that $\sigma' = \sigma$ is the unique solution to equation (29) as long as $M \left[\frac{\partial s_b^{-1}}{\partial \sigma'} - \frac{\partial s_a^{-1}}{\partial \sigma'}\right]\bigg|_{(\sigma', r')=(\sigma, 0)} \neq 0$. This condition can be re-written as:

$$\left[\frac{\partial s_{a(1,2)}^{-1}}{\partial \sigma'} - \frac{\partial s_{a1}^{-1}}{\partial \sigma'} - \frac{\partial s_{a2}^{-1}}{\partial \sigma'}\right]\bigg|_{(\sigma', r')=(\sigma, 0)} \neq \left[\frac{\partial s_{b(1,2)}^{-1}}{\partial \sigma'} - \frac{\partial s_{b1}^{-1}}{\partial \sigma'} - \frac{\partial s_{b2}^{-1}}{\partial \sigma'}\right]\bigg|_{(\sigma', r')=(\sigma, 0)}, \tag{30}$$

or equivalently as:

$$\frac{\partial \Gamma(\mathscr{s}_a; \sigma', r')}{\partial \sigma'}\bigg|_{(\sigma', r')=(\sigma, 0)} \neq \frac{\partial \Gamma(\mathscr{s}_b; \sigma', r')}{\partial \sigma'}\bigg|_{(\sigma', r')=(\sigma, 0)}. \tag{31}$$

Condition (31) makes clear that, in order to achieve identification, the derivative of the *recovered* demand synergies at the true parameters $(\sigma, 0)$ should be different when evaluated at $\mathscr{s}_a$ and at $\mathscr{s}_b$. To the very minimum, condition (31) requires some variation across markets, in the sense

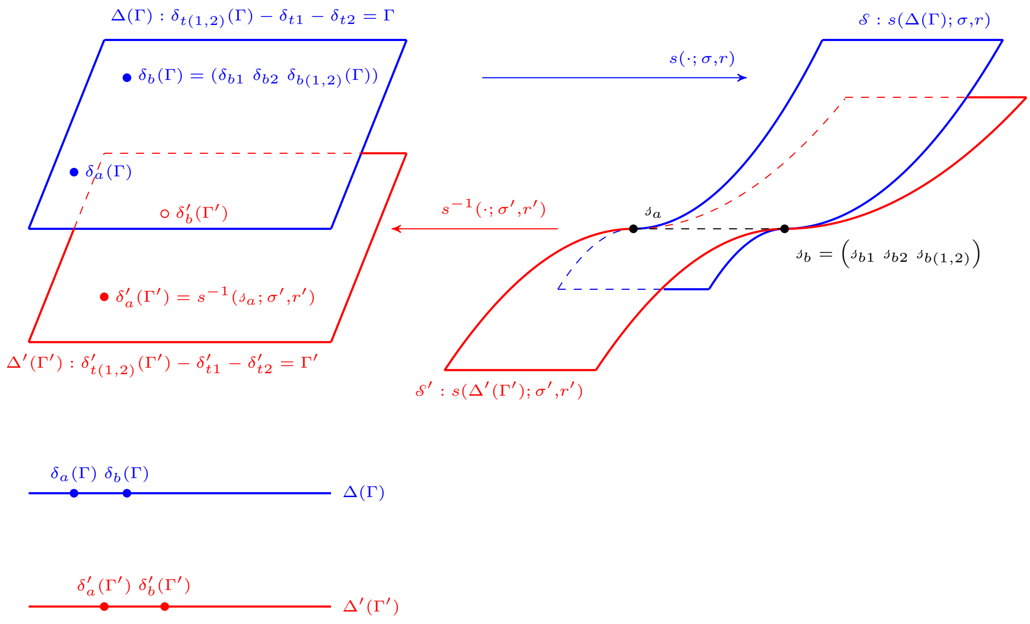


Figure 2: An example of lack of identification

of $\mathscr{s}_a \neq \mathscr{s}_b$. More broadly, given the stark non-linearity of $\frac{\partial \Gamma(\mathscr{s}'; \sigma', r')}{\partial \sigma'}\Big|_{(\sigma', r')=(\sigma, 0)}$, the model will typically be identified whenever $\mathscr{s}_a \neq \mathscr{s}_b$.

Example 2. Suppose there are two markets $T = 2$, $t \in \{a, b\}$, and that both σ and r are unknown. The true structural parameters are $(\delta_{a1}, \delta_{a2}, \delta_{b1}, \delta_{b2}, \Gamma, \sigma, r)$ and the observed market shares are $(\mathscr{s}_a, \mathscr{s}_b)$. System (28) simplifies to the following equation:

$$M \left[\frac{\partial s_b^{-1}}{\partial(\sigma', r')} - \frac{\partial s_a^{-1}}{\partial(\sigma', r')} \right] \Bigg|_{(\sigma', r')=(\sigma, r)} (\sigma' - \sigma, r' - r)^{\mathrm{T}} = 0. \tag{32}$$

Note that $M \left[\frac{\partial s_b^{-1}}{\partial(\sigma', r')} - \frac{\partial s_a^{-1}}{\partial(\sigma', r')} \right] \Bigg|_{(\sigma', r')=(\sigma, r)}$ is of size 1×2 and therefore not of full column rank. It then follows that any solution to equation (32) cannot be unique: in a neighborhood of (σ, r), there exist infinitely many (σ', r') such that equation (32) holds.

Figure 2 provides some visual intuition about the lack of identification in this example. On the left part of the Figure, the true $\delta_a(\Gamma)$ and $\delta_b(\Gamma)$ lie on the plane $\Delta(\Gamma)$ depicted in blue, which represents the set of $\delta(\Gamma)$'s that satisfy the constraints from system (24) evaluated at the true demand synergy Γ. These constraints pin down one of the three coordinates of each $\delta(\Gamma) \in \Delta(\Gamma)$, $\delta(\Gamma) = (\delta_1, \delta_2, \delta_1 + \delta_2 + \Gamma)$. On the right part of the Figure, the observed market shares $\mathscr{s}_a$ and $\mathscr{s}_b$ lie on the manifold $\mathscr{S}$ in blue, which displays all the possible market share values consistent with $s(\cdot; \sigma, r)$ and the true demand synergy Γ. However, because equation (32) has

multiple (σ', r') solutions, $\mathscr{s}_a$ and $\mathscr{s}_b$ do not uniquely belong to $\mathcal{S}$. As shown in the right part of the Figure in red, for any solution to equation (32), (σ', r'), also the corresponding manifold $\mathcal{S}'$ will be consistent with $\mathscr{s}_a$ and $\mathscr{s}_b$. In turn, for given $\mathscr{s}_a$ and $\mathscr{s}_b$, the inverse market share function, $s^{-1}(\cdot;\sigma', r')$, will map respectively to $\delta'_a(\Gamma')$ and to $\delta'_b(\Gamma') \in \Delta'(\Gamma') = s^{-1}(\mathcal{S}';\sigma', r')$ as depicted in red on the left part of the Figure. In other words, there exists $(\delta'_{a1}, \delta'_{a2}, \delta'_{b1}, \delta'_{b2}, \Gamma', \sigma', r') \neq (\delta_{a1}, \delta_{a2}, \delta_{b1}, \delta_{b2}, \Gamma, \sigma, r)$ which also solves system (24) and the model is not identified.

Example 3. Imagine a situation similar to Example 2 but with information on one additional market, so that $T = 3$, $t \in \{a, b, c\}$. The structural parameters are $(\delta_{a1}, \delta_{a2}, \delta_{b1}, \delta_{b2}, \delta_{c1}, \delta_{c2}, \Gamma, \sigma, r)$ and the observed market shares are $(\mathscr{s}_a, \mathscr{s}_b, \mathscr{s}_c)$. System (28) simplifies to:

$$\begin{aligned} &M\left[\frac{\partial s_b^{-1}}{\partial(\sigma', r')} - \frac{\partial s_a^{-1}}{\partial(\sigma', r')}\right]_{(\sigma', r')=(\sigma, r)} (\sigma' - \sigma, r' - r)^{\mathrm{T}} = 0 \\ &M\left[\frac{\partial s_c^{-1}}{\partial(\sigma', r')} - \frac{\partial s_a^{-1}}{\partial(\sigma', r')}\right]_{(\sigma', r')=(\sigma, r)} (\sigma' - \sigma, r' - r)^{\mathrm{T}} = 0. \end{aligned} \tag{33}$$

Note that (σ, r) is the unique solution to linear system (33) and the model is identified as long as the 2×2 matrix

$$\left[\begin{array}{c} M\left(\frac{\partial s_b^{-1}}{\partial(\sigma', r')} - \frac{\partial s_a^{-1}}{\partial(\sigma', r')}\right) \\ M\left(\frac{\partial s_c^{-1}}{\partial(\sigma', r')} - \frac{\partial s_a^{-1}}{\partial(\sigma', r')}\right) \end{array}\right]_{(\sigma', r')=(\sigma, r)} \tag{34}$$

is of full column rank. In Example 2, the corresponding matrix in equation (32) was of size 1×2 and therefore not of full column rank. By adding one observation, $\mathscr{s}_c$, one obtains an additional moment restriction (i.e., an additional row to the matrix) and consequently the possibility of full column rank of matrix (34). The full column rank condition for the 2×2 matrix (34) generalizes identification condition (31) from Example 1.

Figure 3 provides some visual intuition about how the additional observations on market c, $\mathscr{s}_c$, allow for the possibility of identification in this example (as opposed to the lack of identification in Example 2). The main content of Figure 3 is similar to that of Figure 2, with the exception of the additional $\delta_c(\Gamma) \in \Delta(\Gamma)$ and the corresponding $\mathscr{s}_c \in \mathcal{S}$. Differently from Example 2, the additional $\mathscr{s}_c$ and the full column rank of (34) guarantee that there is no manifold $\mathcal{S}'$ other than $\mathcal{S} = s(\Delta(\Gamma);\sigma, r)$ that simultaneously contains $\mathscr{s}_a$, $\mathscr{s}_b$, and $\mathscr{s}_c$. In turn, for any $(\sigma', r') \neq (\sigma, r)$, the inverse market share function, $s^{-1}(\cdot;\sigma', r')$, will not simultaneously map $\mathscr{s}_a$, $\mathscr{s}_b$, and $\mathscr{s}_c$ onto the corresponding plane $\Delta'(\Gamma')$. This is depicted in the left and the lower-left parts of Figure 3, where (in red) $\delta'_a(\Gamma')$ and $\delta'_b(\Gamma')$ lie on $\Delta'(\Gamma')$, while (in black) $\tilde{\delta}'_c$ does not. As a consequence, $(\delta_{a1}, \delta_{a2}, \delta_{b1}, \delta_{b2}, \delta_{c1}, \delta_{c2}, \Gamma, \sigma, r)$ is the unique solution to system (24) and the model is identified.

These three examples highlight two general points about the identification of model (3). First, as condition (34) illustrates, in our framework the task of recovering the full set of

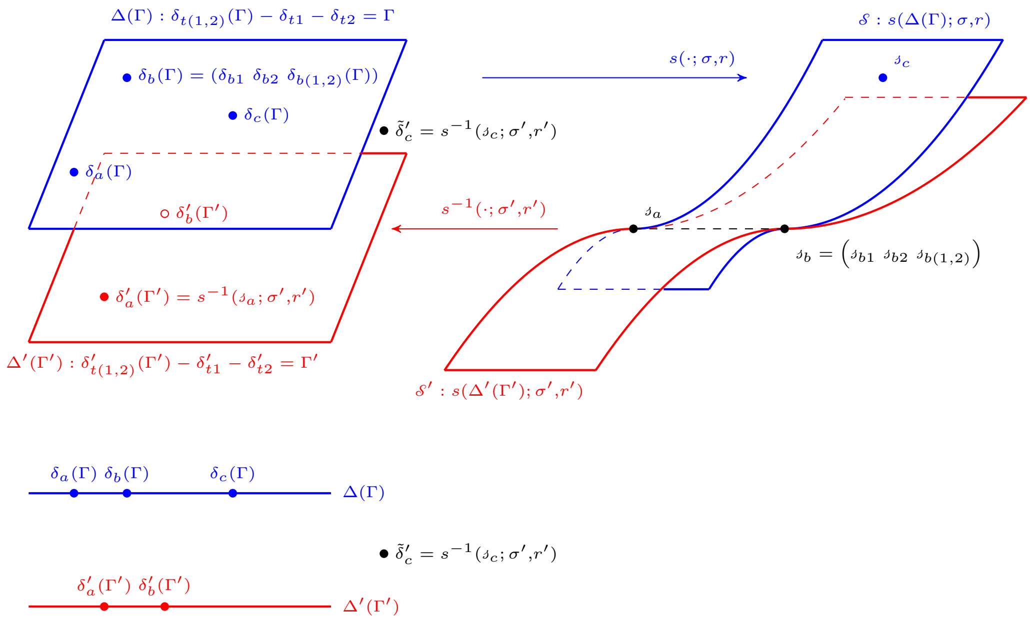


Figure 3: An example of identification

structural parameters reduces to that of identifying the parameters of the distribution of random coefficients Σ. This directly follows from two features of system (24): the invertibility of the market share function $s_t(\cdot; \sigma', r')$ and the common *average* demand synergy parameter Γ across markets in the moment restrictions $\delta_{t(1,2)}(\Gamma) - \delta_{t1} - \delta_{t2} = \Gamma$, $t = 1, ..., T$.[44] Second, as system (28) illustrates, whenever the dimension of Σ does not depend on the number of markets T, adding markets to the dataset will help identification. Identification requires matrix (34) to be of full column rank. In Example 2 the number of markets (i.e., number of rows plus one) is smaller than the dimension of Σ (i.e., number of columns), and identification can hardly be achieved. Differently, by adding one market to the dataset, matrix (34) in Example 3 has as many rows as columns and the model can be identified on the basis of the full column rank condition. Similarly, in Example 1 identification can be achieved because, even though there are only two markets, Σ only contains one parameter, σ, rather than two, σ and r.

7.3 Proof of Lemma 1

To prove the first statement, we show that given a distribution function for β_{it}, $F(\cdot; \Sigma')$, there exists a unique $\delta'_t \in \mathbb{R}^{C_{t1}}$ for $t = 1, ..., T$ that solves $s_t(\delta'_t; \Sigma') = s_t$. This is equivalent to showing that given $F(\cdot; \Sigma')$, the market share function $s_t(\cdot; F)$ is invertible for $t = 1, ..., T$. Because our

[44]Importantly, remember that the demand synergies Γ_i, $i = 1, ..., I$, are heterogeneous across the I individuals and that only their averages Γ_t are constrained to be common across markets, so that $\Gamma_t = \Gamma$, $t = 1, ..., T$.

arguments do not depend on whether F is parametric or non-parametric, in what follows we denote $F(\cdot;\Sigma')$ by F.

Given a distribution F, for market $t = 1, ..., T$, define the Jacobian matrix of the market share function $s_t(\cdot;F)$ from (3) by:

$$\mathbb{J}_t(\delta_t';F) = \frac{\partial s_t}{\partial \delta_t'}(\delta_t';F) = \left(\frac{\partial s_{t\mathbf{b}}}{\partial \delta_{t\mathbf{b}'}'}(\delta_t';F)\right)_{\mathbf{b},\mathbf{b}'\in\mathbf{C}_{t1}}. \tag{35}$$

Corollary 2 from Berry, Gandhi, and Haile (2013) gives sufficient conditions for the invertibility of market share functions that are differentiable. We now verify that the market share function (3) satisfies the two sufficient conditions of Corollary 2 from Berry, Gandhi, and Haile (2013): (a) weak substitutes (Assumption 2 in Berry, Gandhi, and Haile (2013)) and (b) non-singularity of the Jacobian matrix $\mathbb{J}_t(\delta_t';F)$. We first compute $\mathbb{J}_t(\delta_t';F)$ for $\mathbf{b},\mathbf{b}'\in\mathbf{C}_{t1}, \mathbf{b}\neq\mathbf{b}'$:

$$\begin{aligned}\frac{\partial s_{t\mathbf{b}}}{\partial \delta_{t\mathbf{b}}'}(\delta_t';F) &= \int s_{it\mathbf{b}}(\delta_t';\beta_{it})(1-s_{it\mathbf{b}}(\delta_t';\beta_{it}))dF(\beta_{it})\\ \frac{\partial s_{t\mathbf{b}}}{\partial \delta_{t\mathbf{b}'}'}(\delta_t';F) &= -\int s_{it\mathbf{b}}(\delta_t';\beta_{it})s_{it\mathbf{b}'}(\delta_t';\beta_{it})dF(\beta_{it}).\end{aligned} \tag{36}$$

As discussed in Berry, Gandhi, and Haile (2013), the weak substitutes condition does not rule out complementarities in a discrete choice model in which alternatives are defined as bundles, as in our demand model (3) and the model by Gentzkow (2007). In practice, the weak substitutes condition requires that for all $t = 1, ..., T$ and $\mathbf{b}\in\mathbf{C}_{t1}$, $s_{t\mathbf{b}}(\delta_t';F)$ be weakly decreasing in $\delta_{t\mathbf{b}'}'$ for any $\mathbf{b}'\neq\mathbf{b},\mathbf{b}'\in\mathbf{C}_{t1}$. This is immediate from the second equation in (36). In what follows, we verify that $\mathbb{J}_t(\delta_t';F)$ is non-singular.

Define the $C_{t1}\times 1$ vector $s_{it}(\delta_t';\beta_{it}) = (s_{it\mathbf{b}}(\delta_t';\beta_{it}))_{\mathbf{b}\in\mathbf{C}_{t1}}$. By using (36), we can re-write $\mathbb{J}_t(\delta_t';F)$ as:

$$\mathbb{J}_t(\delta_t';F) = \int \left[\mathrm{Diag}(s_{it}(\delta_t';\beta_{it})) - s_{it}(\delta_t';\beta_{it})s_{it}(\delta_t';\beta_{it})^{\mathrm{T}}\right] dF(\beta_{it}), \tag{37}$$

where $\mathrm{Diag}(s_{it}(\delta_t';\beta_{it}))$ is a diagonal matrix with the elements of $s_{it}(\delta_t';\beta_{it})$ on the main diagonal. We first show that the symmetric matrix $\mathrm{Diag}(s_{it}(\delta_t';\beta_{it})) - s_{it}(\delta_t';\beta_{it})s_{it}(\delta_t';\beta_{it})^{\mathrm{T}}$ is positive-definite. This is equivalent to showing that its eigenvalues are all positive. Note that every element of $s_{it}(\delta_t';\beta_{it})$ is strictly positive and that their sum is strictly less than one:

$$\begin{aligned}&s_{it\mathbf{b}}(\delta_t';\beta_{it}) > 0,\\ &\sum_{\mathbf{b}\in\mathbf{C}_{t1}} s_{it\mathbf{b}}(\delta_t';\beta_{it}) < 1.\end{aligned}$$

Denote any of the eigenvalues of $\mathrm{Diag}(s_{it}(\delta_t';\beta_{it})) - s_{it}(\delta_t';\beta_{it})s_{it}(\delta_t';\beta_{it})^{\mathrm{T}}$ by λ and its correspond-

ing (non-degenerate) eigenvector by $\boldsymbol{x}$. Without loss of generality, suppose that the maximal element of vector $\boldsymbol{x}$ in absolute value is its first element $x_1 \neq 0$:

$$|x_1| \geq |x_{\mathbf{b}}| \text{ for any } \mathbf{b} \in \mathbf{C}_{t1}.$$

Then, we have:

$$\begin{aligned}
&\left[\mathrm{Diag}(s_{it}(\delta_t'; \beta_{it})) - s_{it}(\delta_t'; \beta_{it}) s_{it}(\delta_t'; \beta_{it})^{\mathrm{T}}\right] \boldsymbol{x} = \lambda \boldsymbol{x} \\
\Longrightarrow\ & s_{it\mathbf{b}}(\delta_t'; \beta_{it}) x_{\mathbf{b}} - s_{it\mathbf{b}}(\delta_t'; \beta_{it}) \sum_{\mathbf{b}' \in \mathbf{C}_{t1}} s_{it\mathbf{b}'}(\delta_t'; \beta_{it}) x_{\mathbf{b}'} = \lambda x_{\mathbf{b}}, \text{ for all } \mathbf{b} \in \mathbf{C}_{t1} \\
\Longrightarrow\ & s_{it1}(\delta_t'; \beta_{it}) x_1 - s_{it1}(\delta_t'; \beta_{it}) \sum_{\mathbf{b}' \in \mathbf{C}_{t1}} s_{it\mathbf{b}'}(\delta_t'; \beta_{it}) x_{\mathbf{b}'} = \lambda x_1 \\
\Longrightarrow\ & \lambda = s_{it1}(\delta_t'; \beta_{it}) \left(1 - \frac{\sum_{\mathbf{b}' \in \mathbf{C}_{t1}} s_{it\mathbf{b}'}(\delta_t'; \beta_{it}) x_{\mathbf{b}'}}{x_1}\right) \\
& \geq s_{it1}(\delta_t'; \beta_{it}) \left(1 - \left|\frac{\sum_{\mathbf{b}' \in \mathbf{C}_{t1}} s_{it\mathbf{b}'}(\delta_t'; \beta_{it}) x_{\mathbf{b}'}}{x_1}\right|\right) \\
& \geq s_{it1}(\delta_t'; \beta_{it}) \left(1 - \frac{\sum_{\mathbf{b}' \in \mathbf{C}_{t1}} s_{it\mathbf{b}'}(\delta_t'; \beta_{it}) |x_{\mathbf{b}'}|}{|x_1|}\right) \\
& \geq s_{it1}(\delta_t'; \beta_{it}) \left(1 - \sum_{\mathbf{b}' \in \mathbf{C}_{t1}} s_{it\mathbf{b}'}(\delta_t'; \beta_{it})\right) \\
& > 0.
\end{aligned}$$

Any eigenvalue of $\mathrm{Diag}(s_{it}(\delta_t'; \beta_{it})) - s_{it}(\delta_t'; \beta_{it}) s_{it}(\delta_t'; \beta_{it})^{\mathrm{T}}$ is thus strictly positive: for any $v \in \mathbb{R}^{C_{t1}}$,

$$v^{\mathrm{T}}[\mathrm{Diag}(s_{it}(\delta_t'; \beta_{it})) - s_{it}(\delta_t'; \beta_{it}) s_{it}(\delta_t'; \beta_{it})^{\mathrm{T}}] v > 0.$$

As a consequence,

$$\begin{aligned}
v^{\mathrm{T}} \mathbb{J}_t(\delta_t'; F) v &= \int v^{\mathrm{T}}[\mathrm{Diag}(s_{it}(\delta_t'; \beta_{it})) - s_{it}(\delta_t'; \beta_{it}) s_{it}(\delta_t'; \beta_{it})^{\mathrm{T}}] v dF(\beta_{it}) \\
&> 0.
\end{aligned}$$

Thus, given F, for any $\delta_t' \in \mathbb{R}^{C_{t1}}$, $\mathbb{J}_t(\delta_t'; F)$ is positive-definite and non-singular. Because both conditions (a) and (b) of Corollary 2 by Berry, Gandhi, and Haile (2013) are satisfied, then the market share function $s_t(\delta_t'; F)$ is invertible with respect to δ_t', for $t = 1, ..., T$. This completes the proof of the first statement.

According to Assumption 1, the density function $\frac{dF(\beta_{it};\Sigma')}{d\beta_{it}}$ is continuously differentiable with respect to Σ'. As a consequence, $s_t(\delta_t'; \Sigma') - \mathcal{s}_t'$ is continuously differentiable with respect to $(\delta_t', \mathcal{s}_t', \Sigma')$. As we showed above, the Jacobian matrix $\frac{\partial[s_t(\delta_t';\Sigma') - \mathcal{s}_t']}{\partial \delta_t'}\Big|_{(\delta_t', \mathcal{s}_t', \Sigma') = (\delta_t, \mathcal{s}_t, \Sigma)} = \mathbb{J}_t(\delta_t; F(\cdot; \Sigma))$ is invertible. Then, according to the Implicit Function Theorem, in a neigh-

bourhood of $(\delta_t, \mathscr{s}_t, \Sigma)$, for any $(\mathscr{s}_t', \Sigma')$ there exists a unique δ_t' such that $s_t(\delta_t'; \Sigma') = \mathscr{s}_t'$ and $s_t^{-1}(\mathscr{s}_t'; \Sigma') = \delta_t'$ is continuously differentiable with respect to $(\mathscr{s}_t', \Sigma')$. This completes the proof of the second statement.

7.4 Proof of Theorem 1

The first part of the proof consists of the following Lemma. Afterwards, we rely on this Lemma to prove Theorem 1 by contradiction.

Lemma 4. *If Assumptions 1 and 2 hold, and the Jacobian matrix* $\frac{\partial m(\Sigma')}{\partial \Sigma'}\big|_{\Sigma'=\Sigma}$ *is of full column rank, then* Σ *is locally uniquely determined by moment conditions (9).*

Proof. The differentiability of moment conditions (9) with respect to Σ' follows from the second statement of Lemma 1. It then suffices to show that the true Σ is the unique local solution to $m(\Sigma') = 0$. From the definition of model (3), $m(\Sigma) = 0$. We prove the result by contradiction.

Suppose that Σ is not the unique local solution to $m(\Sigma') = 0$. As a consequence, there exists a sequence of Σ_N such that $\Sigma_N \to \Sigma$ as $N \to \infty$, and $m(\Sigma_N) = 0$. Because $m(\Sigma')$ is continuously differentiable in a neighbourhood of $\Sigma' = \Sigma$, by applying a first-order Taylor expansion, we have:

$$\begin{aligned} m(\Sigma_N) &= m(\Sigma) + \frac{\partial m(\Sigma')}{\partial \Sigma'}\bigg|_{\Sigma'=\Sigma} (\Sigma_N - \Sigma) + o(|\Sigma_N - \Sigma|), \\ \frac{\partial m(\Sigma')}{\partial \Sigma'}\bigg|_{\Sigma'=\Sigma} \frac{\Sigma_N - \Sigma}{|\Sigma_N - \Sigma|} &= -\frac{o(|\Sigma_N - \Sigma|)}{|\Sigma_N - \Sigma|}, \end{aligned} \tag{38}$$

where $o(|\Sigma_N - \Sigma|)$ is such that $\lim_{N\to\infty} \frac{o(|\Sigma_N - \Sigma|)}{|\Sigma_N - \Sigma|} = 0$. Note that $\frac{\Sigma_N - \Sigma}{|\Sigma_N - \Sigma|}$ belongs to the unit sphere in $\mathbb{R}^P$, which is compact. Then, there exists a subsequence $\left\{\frac{\Sigma_{N_\ell} - \Sigma}{|\Sigma_{N_\ell} - \Sigma|}\right\}$ and $v \in \mathbb{R}^P$ with $|v| = 1$, such that $\frac{\Sigma_{N_\ell} - \Sigma}{|\Sigma_{N_\ell} - \Sigma|} \to v$. By applying the second equation of (38) to the subsequence $\left\{\frac{\Sigma_{N_\ell} - \Sigma}{|\Sigma_{N_\ell} - \Sigma|}\right\}$, and by combining $\Sigma_{N_\ell} \to \Sigma$ and the continuous differentiability of $m(\cdot)$ in a neighborhood of Σ, we obtain $\frac{\partial m(\Sigma')}{\partial \Sigma'}\big|_{\Sigma'=\Sigma} v = 0$. Because $\frac{\partial m(\Sigma')}{\partial \Sigma'}\big|_{\Sigma'=\Sigma}$ is of full column rank, any vector $x \in \mathbb{R}^P$ that satisfies $\frac{\partial m(\Sigma')}{\partial \Sigma'}\big|_{\Sigma'=\Sigma} x = 0$ must be zero. Then $v = 0$, which contradicts the fact that $|v| = 1$. As a consequence, Σ is the unique local solution to $m(\Sigma') = 0$. □

We now prove Theorem 1 by contradiction. Suppose that model (3) is not locally identified: there exists a sequence of solutions to system (4), $(\delta^N_{1\mathbf{J}_1}, ..., \delta^N_{T\mathbf{J}_T}, \Gamma^N, \Sigma^N) \neq (\delta_{1\mathbf{J}_1}, ..., \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$ for any N, such that $(\delta^N_{1\mathbf{J}_1}, ..., \delta^N_{T\mathbf{J}_T}, \Gamma^N, \Sigma^N) \to (\delta_{1\mathbf{J}_1}, ..., \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$ as $N \to \infty$. Applying (6)

and (7) to each element of the sequence, one obtains:

$$\begin{aligned}
&\delta_{t\mathbf{b}}^{N}(\Gamma_{\mathbf{b}}^{N}) = s_{t\mathbf{b}}^{-1}(\jmath_t;\Sigma^N),\\
&\delta_{tj}^{N} = s_{tj}^{-1}(\jmath_t;\Sigma^N),\ j\in\mathbf{b},\\
&\Gamma_{\mathbf{b}}^{N} = s_{t\mathbf{b}}^{-1}(\jmath_t;\Sigma^N) - \sum_{j\in\mathbf{b}} s_{tj}^{-1}(\jmath_t;\Sigma^N).
\end{aligned} \tag{39}$$

Then, by constructing moment conditions (9) for each element of the sequence, we have $m(\Sigma')|_{\Sigma'=\Sigma^N} = 0$ for any N. Because the Jacobian matrix $\frac{\partial m(\Sigma')}{\partial \Sigma'}\big|_{\Sigma'=\Sigma}$ is of full column rank, according to Lemma 4, Σ is uniquely locally determined by moment conditions (9). Hence, there exists N_0 such that for all $N \geq N_0$, $\Sigma^N = \Sigma$. Because of the third equation of (39), for all $N \geq N_0$, $\Gamma_{\mathbf{b}}^N = \Gamma_{\mathbf{b}}$. Moreover, because of the first two equations of (39), $\delta_{t\mathbf{b}}^N = \delta_{t\mathbf{b}}$, for all $N \geq N_0$, $t = 1, ..., T$ and $\mathbf{b} \in \mathbf{C}_{t1}$. As a consequence, $(\delta_{1\mathbf{J}_1}^N, ..., \delta_{T\mathbf{J}_T}^N, \Gamma^N, \Sigma^N) = (\delta_{1\mathbf{J}_1}, ..., \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$ for all $N \geq N_0$, which contradicts $(\delta_{1\mathbf{J}_1}^N, ..., \delta_{T\mathbf{J}_T}^N, \Gamma^N, \Sigma^N) \neq (\delta_{1\mathbf{J}_1}, ..., \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$ for any N.

7.5 Proof of Rank Regularity Property

According to Assumption 1, Υ is a topological space in $\mathbb{R}^P$. Moreover, $\frac{\partial m(\Sigma')}{\partial \Sigma'}$ is continuous with respect to $\Sigma' \in \Upsilon$. According to Property 4 from Lewis (2009), the set of rank regular points for $\frac{\partial m(\Sigma')}{\partial \Sigma'}$ is open and dense in Υ. This completes the proof.

7.6 Proof of Theorem 2

Sufficiency follows from Theorem 1. Next, we prove necessity. To simplify notation, denote the number of moment conditions $\sum_{\mathbf{b}\in\mathbf{C}_2,|\mathbf{T_b}|\geq 2}(|\mathbf{T_b}|-1)$ by Q and the rank of $\frac{\partial m(\Sigma')}{\partial \Sigma'})\big|_{\Sigma'=\Sigma}$ by r. According to Definition 17, there exists a neighbourhood of the true Σ, U, such that $\mathrm{rank}(\frac{m(\Sigma')}{\partial \Sigma'}) = \mathrm{rank}(\frac{m(\Sigma')}{\partial \Sigma'})\big|_{\Sigma'=\Sigma} = r$ for each $\Sigma' \in U$. By applying the Constant Rank Theorem at $\Sigma' = \Sigma$, there are open sets $U_1, U_2 \subset \mathbb{R}^P$ and $U_3 \subset \mathbb{R}^Q$ and diffeomorphisms $\phi : U_1 \to U_2$, $\psi : U_3 \to U_3$ such that $\Sigma \in U_1$ and $\psi \circ m \circ \phi^{-1}(x') = (x'_1, \dots, x'_r, 0, \dots, 0)$ for all $x' \in U_2$.

Define $x = (x_1, ..., x_P) = \phi(\Sigma)$ and a sequence $\{x^N = (x_1^N, ..., x_P^N)\}$ such that $x_\ell^N = x_\ell$, for $\ell = 1, ..., r$ and $x_\ell^N = x_\ell + \frac{1}{N}$, for N large enough so that $x^N \neq x$ and $x^N \in U_2$. Note that

$$\begin{aligned}
\psi \circ m \circ \phi^{-1}(x) &= (x_1, ...x_r, 0, \dots, 0)\\
&= (x_1^N, ...x_r^N, 0, \dots, 0)\\
&= \psi \circ m \circ \phi^{-1}(x^N)
\end{aligned} \tag{40}$$

and that

$$\begin{aligned}\psi\circ m\circ\phi^{-1}(x) &= \psi\circ m\circ\phi^{-1}(\phi(\Sigma))\\ &= \psi\circ m(\Sigma)\\ &= \psi(0).\end{aligned}\tag{41}$$

As a consequence, $\psi\circ m\circ\phi^{-1}(x^N) = \psi(0)$. Because ψ is a diffeomorphism, we obtain $m(\phi^{-1}(x^N)) = 0$. Because ϕ and its inverse ϕ^{-1} are diffeomorphisms and $x \neq x^N \to x = \phi(\Sigma)$ as $N\to\infty$, we construct a sequence $\Sigma^N = \phi^{-1}(x^N) \to \phi^{-1}(x) = \Sigma$ with $\Sigma^N \neq \Sigma$ such that $m(\Sigma^N) = 0$ for each N. According to equations (39) from the proof of Theorem 1, given Σ^N, we can construct a $(\delta^N_{1\mathbf{J}_1}, ..., \delta^N_{T\mathbf{J}_T}, \Gamma^N, \Sigma^N)$ such that it is a solution to (4). Consequently, model (3) is not locally identified.

7.7 Proof of Corollary 1

Because $\frac{\partial m(\Sigma')}{\partial\Sigma'}\big|_{\Sigma'=\Sigma}$ is of full row rank, then the positive definite matrix $\left[\frac{\partial m(\Sigma')}{\partial\Sigma'}\right]\left[\frac{\partial m(\Sigma')}{\partial\Sigma'}\right]^{\mathrm{T}}\big|_{\Sigma'=\Sigma}$ is not singular and its determinant $\mathrm{Det}\left(\left[\frac{\partial m(\Sigma')}{\partial\Sigma'}\right]\left[\frac{\partial m(\Sigma')}{\partial\Sigma'}\right]^{\mathrm{T}}\big|_{\Sigma'=\Sigma}\right)$ is positive. Moreover, since $\frac{\partial m(\Sigma')}{\partial\Sigma'}$ is continuous with respect to Σ', $\mathrm{Det}\left(\left[\frac{\partial m(\Sigma')}{\partial\Sigma'}\right]\left[\frac{\partial m(\Sigma')}{\partial\Sigma'}\right]^{\mathrm{T}}\right)$ is also continuous with respect to Σ' and is positive in a neighbourhood of $\Sigma' = \Sigma$. This implies that $\frac{\partial m(\Sigma')}{\partial\Sigma'}$ is of full row rank in the neighbourhood of $\Sigma' = \Sigma$, and its rank, $\mathrm{rank}\left(\frac{\partial m(\Sigma')}{\partial\Sigma'}\right)$, is constant and equal to the number of rows in $\frac{\partial m(\Sigma')}{\partial\Sigma'}\big|_{\Sigma'=\Sigma}$. Consequently, Σ is rank regular for $\frac{\partial m(\Sigma')}{\partial\Sigma'}$. Note that the number of rows in $\frac{\partial m(\Sigma')}{\partial\Sigma'}\big|_{\Sigma'=\Sigma}$ is equal to the number of moment conditions $\sum_{\mathbf{b}\in\mathbf{C}_2,|\mathbf{T_b}|\geq 2}(|\mathbf{T_b}|-1)$ and it is strictly smaller than the dimension of Σ. The latter is equal to the number of columns in $\frac{\partial m(\Sigma')}{\partial\Sigma'}\big|_{\Sigma'=\Sigma}$. Then, $\frac{\partial m(\Sigma')}{\partial\Sigma'}\big|_{\Sigma'=\Sigma}$ is not of full column rank. According to Theorem 2, model (3) is not locally identified.

7.8 Proof of Lemma 2

Note that the solution set of system (9) is a subset of that of $m(\Sigma';\mathbf{T}_0) = 0$ whenever $\mathbf{T}_0$ is a subset of $\mathbf{T}$. Then, it suffices to prove that the solution set of $m(\Sigma';\mathbf{T}_0) = 0$ in Θ_Σ is finite.

We prove the result by contradiction. Denote the solution set of $m(\Sigma';\mathbf{T}_0) = 0$ in Θ_Σ by $S(\Sigma')$. Suppose that $S(\Sigma')$ contains infinitely many points. Because $S(\Sigma')$ is a closed subset of the compact set Θ_Σ, $S(\Sigma')$ is itself compact. Consequently, due to the infinity of points in $S(\Sigma')$, there exists an accumulation point $\Sigma'_0 \in S(\Sigma')$: in any neighbourhood of Σ'_0, we can find another point $\Sigma''_0 \in S(\Sigma')$, i.e. another solution to $m(\Sigma';\mathbf{T}_0) = 0$. Due to Assumption 3, we know that at $\Sigma'_0 \in S(\Sigma')$, the corresponding Jacobian matrix $\frac{\partial m(\Sigma';\mathbf{T}_0)}{\partial\Sigma'}{}_{\Sigma'=\Sigma'_0}$ is of full column rank. Then, $\Sigma' = \Sigma'_0$ is the local unique solution to $m(\Sigma';\mathbf{T}_0) = 0$. This is in contradiction with Σ'_0 being an accumulation point in $S(\Sigma')$.

7.9 Proof of Theorem 3

Remember that

$$\begin{aligned} s_{t\mathbf{b}}(\delta_t; F) &= \int s_{i\mathbf{b}}(\delta_t, \beta_i) dF(\beta_i) \\ &= \int \frac{e^{\delta_{t\mathbf{b}}+\mu_{i\mathbf{b}}(\beta_i)}}{\sum_{\mathbf{b}'\in\mathbf{C}_t} e^{\delta_{t\mathbf{b}'}+\mu_{i\mathbf{b}'}(\beta_i)}} dF(\beta_i). \end{aligned}$$

To prove the real analytic property of the market share function $s_{t\mathbf{b}}(\delta_t; F)$, it suffices to study $\frac{\partial^l s_{i\mathbf{b}}(\delta_t;\beta_i)}{\prod_{\mathbf{b}'\in\mathbf{C}_{t1}} \partial\delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}}$, where l is an integer and $\sum_{\mathbf{b}'\in\mathbf{C}_{t1}} l_{\mathbf{b}'} = l$. We first prove the following Lemma.

Lemma 5. *For any non-negative integer l,*

$$\sup_{\delta_t,\beta_i} \left| \frac{\partial^l s_{i\mathbf{b}}(\delta_t;\beta_i)}{\partial\delta_{i\mathbf{b}}^l} \right| \leq A_l l!, \tag{42}$$

where $A_l = (e-1)^l \sum_{k=0}^{l} \frac{1}{(e-1)^k k!}$.

Proof. Define $a_l = \sup\limits_{\delta_t,\beta_i} \left| \frac{\partial^l s_{i\mathbf{b}}(\delta_t;\beta_i)}{\partial\delta_{t\mathbf{b}}^l} \right|$. Note that:

$$\begin{aligned} e^{\delta_{t\mathbf{b}}+\mu_{i\mathbf{b}}(\beta_i)} &= s_{i\mathbf{b}} \sum_{\mathbf{b}'\in\mathbf{C}_t} e^{\delta_{t\mathbf{b}'}+\mu_{i\mathbf{b}'}(\beta_i)} \\ e^{\delta_{t\mathbf{b}}+\mu_{i\mathbf{b}}(\beta_i)} &= \frac{\partial^l e^{\delta_{t\mathbf{b}}+\mu_{i\mathbf{b}}(\beta_i)}}{\partial\delta_{t\mathbf{b}}^l} \\ &= \frac{\partial^l \left(s_{i\mathbf{b}} \sum_{\mathbf{b}'\in\mathbf{C}_t} e^{\delta_{t\mathbf{b}'}+\mu_{i\mathbf{b}'}(\beta_i)}\right)}{\partial\delta_{t\mathbf{b}}^l} \\ &= \sum_{k=0}^{l} C_l^k \frac{\partial^k s_{i\mathbf{b}}}{\partial\delta_{t\mathbf{b}}^k} \frac{\partial^{l-k} \sum_{\mathbf{b}'\in\mathbf{C}_t} e^{\delta_{t\mathbf{b}'}+\mu_{i\mathbf{b}'}(\beta_i)}}{\partial\delta_{t\mathbf{b}}^{l-k}} \\ &= \frac{\partial^l s_{i\mathbf{b}}}{\partial\delta_{t\mathbf{b}}^l} \sum_{\mathbf{b}'\in\mathbf{C}_t} e^{\delta_{t\mathbf{b}'}+\mu_{i\mathbf{b}'}(\beta_i)} + \sum_{k=0}^{l-1} C_l^k \frac{\partial^k s_{i\mathbf{b}}}{\partial\delta_{t\mathbf{b}}^k} e^{\delta_{t\mathbf{b}}+\mu_{i\mathbf{b}}(\beta_i)}, \\ \frac{\partial^l s_{i\mathbf{b}}}{\partial\delta_{t\mathbf{b}}^l} &= s_{i\mathbf{b}} \left(1 - \sum_{k=0}^{l-1} C_l^k \frac{\partial^k s_{i\mathbf{b}}}{\partial\delta_{t\mathbf{b}}^k}\right), \\ \left|\frac{\partial^l s_{i\mathbf{b}}}{\partial\delta_{t\mathbf{b}}^l}\right| &\leq 1 + \sum_{k=0}^{l-1} C_l^k \left|\frac{\partial^k s_{i\mathbf{b}}}{\partial\delta_{t\mathbf{b}}^k}\right|, \\ a_l &\leq 1 + \sum_{k=0}^{l-1} C_l^k a_k, \\ \frac{a_l}{l!} &\leq \frac{1}{l!} + \sum_{k=0}^{l-1} \frac{a_k}{k!} \frac{1}{(l-k)!}. \end{aligned} \tag{43}$$

We now show that $\frac{a_l}{l!} \leq A_l$ by induction. For $l = 1$, we have $a_1 = \sup_{\delta_t,\beta_i} \left|\frac{\partial s_{i\mathbf{b}}(\delta_t;\beta_i)}{\partial \delta_{t\mathbf{b}}}\right| = \sup_{\delta_t,\beta_i} |s_{i\mathbf{b}}(1 - s_{i\mathbf{b}})| \leq \frac{1}{4} < e = A_1$. Suppose that $\frac{a_k}{k!} \leq A_k$ holds for $k = 1, ..., l-1$. Note that $A_l = \frac{1}{l!} + (e-1)A_{l-1} > A_{l-1}$, for any $l \geq 0$. Then,

$$
\begin{aligned}
\frac{a_l}{l!} &\leq \frac{1}{l!} + \sum_{k=0}^{l-1} \frac{a_k}{k!} \frac{1}{(l-k)!} \\
&\leq \frac{1}{l!} + \sum_{k=0}^{l-1} A_k \frac{1}{(l-k)!} \\
&\leq \frac{1}{l!} + A_{l-1} \sum_{k=0}^{l-1} \frac{1}{(l-k)!} \\
&\leq \frac{1}{l!} + A_{l-1}(e-1) \\
&= A_l.
\end{aligned} \tag{44}
$$

As a consequence, the inequality holds for any $l > 0$ and $a_l = \sup_{\delta_t,\beta_i} \left|\frac{\partial^l s_{i\mathbf{b}}(\delta_t;\beta_i)}{\partial \delta_{t\mathbf{b}}^l}\right| \leq A_l l!$. This completes the proof. □

The next Lemma controls the size of $\frac{\partial^l s_{i\mathbf{b}}(\delta_t;\beta_i)}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}}$.

Lemma 6. *Suppose $C_{t1} \geq 2$. For any $\mathbf{b} \in \mathbf{C}_{t1}$ and $l > 0$,*

$$
\left|\frac{\partial^l s_{i\mathbf{b}}(\delta_t;\beta_i)}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}}\right| \leq [C_{t1}(e-1)]^l \prod_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'}!, \tag{45}
$$

where $\sum_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'} = l$.

Proof. We prove the result by induction. For $l = 1$, the result follows directly from Lemma 5 with $l = 1$. For $l = 2$ and $l_{\mathbf{b}'} = 2$, according to Lemma 5, we have $\left|\frac{\partial^2 s_{i\mathbf{b}}}{\partial \delta_{t\mathbf{b}'}^2}\right| \leq A_2 2!$. For $l = 2$ and $l_{\mathbf{b}'} = l_{\mathbf{b}''} = 1, \mathbf{b}' \neq \mathbf{b}''$:

$$
\begin{aligned}
&e^{\delta_{t\mathbf{b}} + \mu_{i\mathbf{b}}(\beta_i)} = s_{i\mathbf{b}} \sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'} + \mu_{i\mathbf{b}'}(\beta_i)}, \\
&0 = \frac{\partial^2 s_{i\mathbf{b}''}}{\partial \delta_{t\mathbf{b}'} \partial \delta_{t\mathbf{b}''}} \sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'} + \mu_{i\mathbf{b}'}(\beta_i)} + e^{\delta_{t\mathbf{b}'} + \mu_{i\mathbf{b}'}(\beta_i)} \frac{\partial s_{i\mathbf{b}}}{\partial \delta_{t\mathbf{b}''}} + e^{\delta_{t\mathbf{b}''} + \mu_{i\mathbf{b}''}(\beta_i)} \frac{\partial s_{i\mathbf{b}}}{\partial \delta_{t\mathbf{b}'}}, \\
&\frac{\partial^2 s_{i\mathbf{b}}}{\partial \delta_{t\mathbf{b}'} \partial \delta_{t\mathbf{b}''}} = -s_{i\mathbf{b}'} \frac{\partial s_{i\mathbf{b}}}{\partial \delta_{t\mathbf{b}''}} - s_{i\mathbf{b}''} \frac{\partial s_{i\mathbf{b}}}{\partial \delta_{t\mathbf{b}'}},
\end{aligned} \tag{46}
$$

By using $\left|\frac{\partial s_{i\mathbf{b}}}{\partial \delta_{t\mathbf{b}}}\right| \leq \frac{1}{4} < 1$ and $\left|\frac{\partial s_{i\mathbf{b}}}{\partial \delta_{t\mathbf{b}'}}\right| \leq 1$, we have $\left|\frac{\partial^2 s_{i\mathbf{b}}}{\partial \delta_{t\mathbf{b}'} \partial \delta_{t\mathbf{b}}}\right| \leq \left|\frac{\partial s_{i\mathbf{b}}}{\partial \delta_{t\mathbf{b}}}\right| + \left|\frac{\partial s_{i\mathbf{b}}}{\partial \delta_{t\mathbf{b}'}}\right| \leq 2 \leq [C_{t1}(e-1)]^2$.

Note that $A_2 = (e-1)^2(1+\frac{1}{e-1}+\frac{1}{2(e-1)^2}) \leq [C_{t1}(e-1)]^2$ for $C_{t1} \geq 2$. As a consequence, the conclusion holds for $l = 2$ and $\sum\limits_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'} = 2$.

Suppose that for $k = 1, ..., l-1$ the inequality holds for any $\sum\limits_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'} = k$. First, remember that $A_l = (e-1)^l \sum_{k=0}^{l} \frac{1}{(e-1)^k k!}$, as defined in Lemma 5, is smaller than $[C_{t1}(e-1)]^l$ because $C_{t1} \geq 2$. Then, the conclusion holds for any $l > 0$ with $l_{\mathbf{b}'} = l$ and $l_{\mathbf{b}''} = 0$, $\mathbf{b}'' \neq \mathbf{b}'$. It remains to show that the conclusion holds when there exit $\mathbf{b}'$ and $\mathbf{b}''$ such that $l_{\mathbf{b}'} > 0$ and $l_{\mathbf{b}''} > 0$.

By taking $l_{\mathbf{b}}$-th derivatives of both sides of the first equation in (43) with respect to $\delta_{i\mathbf{b}}$, we obtain:

$$\begin{aligned} e^{\delta_{t\mathbf{b}}+\mu_{i\mathbf{b}}(\beta_i)} &= \frac{\partial^{l_{\mathbf{b}}} e^{\delta_{t\mathbf{b}}+\mu_{i\mathbf{b}}(\beta_i)}}{\partial \delta_{t\mathbf{b}}^{l_{\mathbf{b}}}} \\ &= \frac{\partial^{l_{\mathbf{b}}} \left(s_{i\mathbf{b}} \sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'}+\mu_{i\mathbf{b}'}(\beta_i)} \right)}{\partial \delta_{t\mathbf{b}}^{l_{\mathbf{b}}}} \\ &= \sum_{k=0}^{l_{\mathbf{b}}} C_{l_{\mathbf{b}}}^{k} \frac{\partial^k s_{i\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^{k}} \frac{\partial^{l_{\mathbf{b}}-k} \sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'}+\mu_{i\mathbf{b}'}(\beta_i)}}{\partial \delta_{t\mathbf{b}}^{l_{\mathbf{b}}-k}} \\ &= \frac{\partial^{l_{\mathbf{b}}} s_{i\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^{l_{\mathbf{b}}}} \sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'}+\mu_{i\mathbf{b}'}(\beta_i)} + e^{\delta_{t\mathbf{b}}+\mu_{i\mathbf{b}}(\beta_i)} \sum_{k=0}^{l_{\mathbf{b}}-1} C_{l_{\mathbf{b}}}^{k} \frac{\partial^k s_{i\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^{k}}. \end{aligned} \tag{47}$$

Note that, by taking derivatives of both sides of equation (47) with respect to $\delta_{t\mathbf{b}'}$, $\mathbf{b}' \neq \mathbf{b}$, the left hand-side vanishes and we obtain:

$$0 = \frac{\partial^{l_{\mathbf{b}}+l_{\mathbf{b}'}} s_{i\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^{l_{\mathbf{b}}} \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}} \sum_{\mathbf{b}'' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}''}+\mu_{i\mathbf{b}''}(\beta_i)} + e^{\delta_{t\mathbf{b}'}+\mu_{i\mathbf{b}'}(\beta_i)} \sum_{k=0}^{l_{\mathbf{b}'}-1} C_{l'_{\mathbf{b}}}^{k} \frac{\partial^{l_{\mathbf{b}}+k} s_{i\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^{l_{\mathbf{b}}} \partial \delta_{t\mathbf{b}'}^{k}} + e^{\delta_{t\mathbf{b}}+\mu_{i\mathbf{b}}(\beta_i)} \sum_{k=0}^{l_{\mathbf{b}}-1} C_{l_{\mathbf{b}}}^{k} \frac{\partial^{k+l_{\mathbf{b}'}} s_{i\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^{k} \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}}. \tag{48}$$

By taking $l_{\mathbf{b}'}$-th derivatives with respect to $\delta_{t\mathbf{b}'}$, for all $\mathbf{b}' \in \mathbf{C}_{t1}$:

$$\begin{aligned} 0 &= \frac{\partial^{l} s_{i\mathbf{b}}}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}} \sum_{\mathbf{b}'' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}''}+\mu_{i\mathbf{b}''}(\beta_i)} + \sum_{\mathbf{b}' \in \mathbf{C}_{t1}} e^{\delta_{t\mathbf{b}'}+\mu_{i\mathbf{b}'}(\beta_i)} \sum_{k=0}^{l_{\mathbf{b}'}-1} C_{l_{\mathbf{b}'}}^{k} \frac{\partial^{l-l_{\mathbf{b}'}+k} s_{i\mathbf{b}'}}{\partial \delta_{t\mathbf{b}'}^{k} \prod_{\mathbf{b}'' \neq \mathbf{b}'} \partial \delta_{t\mathbf{b}''}^{l_{\mathbf{b}''}}}, \\ \frac{\partial^{l} s_{i\mathbf{b}}}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}} &= - \sum_{\mathbf{b}' \in \mathbf{C}_{t1}} s_{i\mathbf{b}'} \sum_{k=0}^{l_{\mathbf{b}'}-1} C_{l_{\mathbf{b}'}}^{k} \frac{\partial^{l-l_{\mathbf{b}'}+k} s_{i\mathbf{b}'}}{\partial \delta_{t\mathbf{b}'}^{k} \prod_{\mathbf{b}'' \neq \mathbf{b}'} \partial \delta_{t\mathbf{b}''}^{l_{\mathbf{b}''}}}, \\ \frac{\frac{\partial^{l} s_{i\mathbf{b}}}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}}}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'}!} &= - \sum_{\mathbf{b}' \in \mathbf{C}_{t1}} s_{i\mathbf{b}'} \sum_{k=0}^{l_{\mathbf{b}'}-1} \frac{1}{(l_{\mathbf{b}'}-k)!} \frac{\frac{\partial^{l-l_{\mathbf{b}'}+k} s_{i\mathbf{b}'}}{\partial \delta_{t\mathbf{b}'}^{k} \prod_{\mathbf{b}'' \neq \mathbf{b}'} \partial \delta_{t\mathbf{b}''}^{l_{\mathbf{b}''}}}}{k! \prod_{\mathbf{b}'' \neq \mathbf{b}'} l_{\mathbf{b}'}!}, \\ \sup_{\delta_t, \beta_i} \left| \frac{\frac{\partial^{l} s_{i\mathbf{b}}}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}}}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'}!} \right| &\leq \sum_{\mathbf{b}' \in \mathbf{C}_{t1}} \sum_{k=0}^{l_{\mathbf{b}'}-1} \frac{1}{(l_{\mathbf{b}'}-k)!} \sup_{\delta_t, \beta_i} \left| \frac{\frac{\partial^{l-l_{\mathbf{b}'}+k} s_{i\mathbf{b}'}}{\partial \delta_{t\mathbf{b}'}^{k} \prod_{\mathbf{b}'' \neq \mathbf{b}'} \partial \delta_{i\mathbf{b}''}^{l_{\mathbf{b}''}}}}{k! \prod_{\mathbf{b}'' \neq \mathbf{b}'} l_{\mathbf{b}'}!} \right|. \end{aligned} \tag{49}$$

Then, applying the conclusion for any $k = 1, ..., l-1$ on the last equation in (49), we obtain:

$$
\begin{aligned}
\sup_{\delta_t, \beta_i} \left| \frac{\frac{\partial^l s_{i\mathbf{b}}}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}}}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'}!} \right| &\leq \sum_{\mathbf{b}' \in \mathbf{C}_{t1}} \sum_{k=0}^{l_{\mathbf{b}'}-1} \frac{1}{(l_{\mathbf{b}'} - k)!} [C_{t1}(e-1)]^{l - l_{\mathbf{b}'} + k} \\
&= [C_{t1}(e-1)]^l \sum_{\mathbf{b}' \in \mathbf{C}_{t1}} \sum_{k=1}^{l_{\mathbf{b}'}-1} \frac{1}{k!} [C_{t1}(e-1)]^{-k} \\
&\leq [C_{t1}(e-1)]^l C_{t1} (e^{[C_{t1}(e-1)]^{-1}} - 1) \\
&\leq [C_{t1}(e-1)]^l [e^{(e-1)^{-1}} - 1] \\
&< [C_{t1}(e-1)]^l.
\end{aligned}
\tag{50}
$$

Hence, the conclusion holds for $\sum\limits_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'} = l$, and $\sup\limits_{\delta_t, \beta_i} \left| \frac{\partial^l s_{i\mathbf{b}}}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}} \right| \leq [C_{t1}(e-1)]^l \prod_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'}!$ for any $l > 0$ and $\sum\limits_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'} = l$. The proof is completed. □

The size of the l-th derivative of $s_{t\mathbf{b}}(\delta_t; F)$ with respect to δ_t can then be controlled as:

$$
\begin{aligned}
\left| \frac{\partial^l s_{t\mathbf{b}}(\delta_t; F)}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}} \right| &\leq \int \left| \frac{\partial^l s_{i\mathbf{b}}(\delta_t; \beta_i)}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}} \right| dF(\beta_i) \\
&\leq [C_{t1}(e-1)]^l \prod_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'}!
\end{aligned}
\tag{51}
$$

and, consequently, the Taylor expansion of $s_{t\mathbf{b}}(.; F)$ at some δ'_t around δ_t as:

$$
\begin{aligned}
\left| \sum_{L=0}^{\infty} \frac{1}{L!} \left[\sum_{\mathbf{b}' \in \mathbf{C}_{t1}} (\delta'_{t\mathbf{b}'} - \delta_{t\mathbf{b}'}) \frac{\partial}{\partial \delta_{t\mathbf{b}'}} \right]^L s_{t\mathbf{b}}(\delta_t; F) \right| &\leq \sum_{L=0}^{\infty} \frac{1}{L!} d^L \sum_{\sum l_{\mathbf{b}'} = L} \frac{L!}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'}!} \left| \frac{\partial^L s_{t\mathbf{b}}(\delta_t; F)}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}} \right| \\
&\leq \sum_{L=0}^{\infty} d^L C_{t1}^L [C_{t1}(e-1)]^L,
\end{aligned}
\tag{52}
$$

where $d = |\delta'_t - \delta_t|$. Consequently, whenever $d < d^* = \frac{1}{C_{t1}^2 (e-1)}$, the Taylor expansion (52) converges. Finally, by applying Taylor's Theorem to the multivariate function $s_{t\mathbf{b}}(\delta'_t; F)$, we

obtain for any $R > 0$ and uniformly for $|\delta'_t - \delta_t| < d^*$:

$$
\begin{aligned}
&\left| s_{t\mathbf{b}}(\delta'_t; F) - \sum_{L=0}^{R} \frac{1}{L!} \left[\sum_{\mathbf{b}' \in \mathbf{C}_{t1}} (\delta'_{t\mathbf{b}'} - \delta_{t\mathbf{b}'}) \frac{\partial}{\partial \delta_{t\mathbf{b}'}} \right]^L s_{t\mathbf{b}}(\delta_t; F) \right| \\
\leq & d^{R+1} \sum_{\sum l_{\mathbf{b}'} = R+1} \frac{1}{\prod l_{\mathbf{b}'}!} \sup_{|\delta'_t - \delta_t| < d} \left| \frac{\partial^{R+1} s_{t\mathbf{b}}(\delta'_t; F)}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}} \right| \\
\leq & d^{R+1} [C_{t1}(e-1)]^{R+1} C_{t1}^{R+1} \\
\to & 0.
\end{aligned}
$$

In conclusion, the market share function $s_{t\mathbf{b}}(\delta'_t; F)$ is equal to its Taylor expansion and therefore real analytic with respect to δ'_t. This completes the proof.

7.10 Proof of Theorem 4

To ease exposition, we introduce some additional notation. Define the set of matrices $\mathbf{M} = \{M_t : t = 1, ..., T\}$ where each M_t is a matrix of dimension $C_{t2} \times C_{t1}$. Remember that C_{t2} the number of multi-product bundles and C_{t1} is the number of non-empty bundles. M_t is made of two sub-matrices: $M_t = [M_t^1, M_t^2]$. M_t^1 is a matrix of 1's and 0's of dimension $C_{t2} \times J_t$, where the columns represent individual products and the rows represent multi-product bundles. Each row of M_i^1 identifies with 1's the product composition of the corresponding multi-product bundle. Moreover, M_t^2 is the identity matrix of dimension $C_{t2} \times C_{t2}$ (i.e., $M_t^2 = I_{C_{t2} \times C_{t2}}$), with the rows corresponding to multi-product bundles. For example, suppose the choice set (without outside option) of individuals in market t to be $\{1, 2, 3, (1, 2), (1, 3), (2, 3)\}$ and the corresponding average utility vector to be $\delta_t = (\delta_{t1}, \delta_{t2}, \delta_{t3}, \delta_{t(1,2)}, \delta_{t(1,3)}, \delta_{t(2,3)})^{\mathrm{T}}$, with $C_{t1} = 6$ and $C_{t2} = 3$. Then the corresponding M_t is:

$$
M_t = \begin{bmatrix} -1 & -1 & 0 & 1 & 0 & 0 \\ -1 & 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & -1 & 0 & 0 & 1 \end{bmatrix}.
$$

To prove the first point, we use Lemma 1 and re-write Δ_r^{ID} as:

$$
\Delta_r^{\mathrm{ID}} = \{(\delta_{t\mathbf{J}_t})_{t \in \mathbf{T} \setminus \mathbf{T}_0} : \exists t \in \mathbf{T} \setminus \mathbf{T}_0 \text{ such that } M_t s_t^{-1}(s_t(\delta_t(\Gamma^0); \Sigma_0); \Sigma_r) \neq \Gamma^r\}.
$$

On the one hand, when $(\delta^0_{t\mathbf{J}_t})_{t \in \mathbf{T} \setminus \mathbf{T}_0} \in \Delta^{\mathrm{ID}} = \cap_{r=1}^R \Delta_r^{\mathrm{ID}}$, for any $r = 1, .., R$ there exists some market $t^+ \in \mathbf{T} \setminus \mathbf{T}_0$ such that $M_{t^+} s_{t^+}^{-1}(\mathfrak{s}_{t^+}; \Sigma_r) \neq \Gamma^r$. On the other, according to the definition of Σ_r, $m(\Sigma_r; \mathbf{T}_0) = 0$. Then, by construction $m(\Sigma_r; \mathbf{T}_0 \cup \{t^+\}) \neq 0$ and therefore $m(\Sigma_r; \mathbf{T}) \neq 0$

for $r = 1, ..., R$.

Remember that the set of solutions to $m(\Sigma'; \mathbf{T}_0) = 0$ in Θ_Σ is $S(\Sigma) = \{\Sigma_r : r = 0, ..., R\}$. Consequently, the set of solutions to $m(\Sigma'; \mathbf{T}) = 0$ is a subset of $S(\Sigma)$. Given that $m(\Sigma_r; \mathbf{T}) \neq 0$ for $r = 1, ..., R$, and that $m(\Sigma_0; \mathbf{T}) = 0$, $\Sigma' = \Sigma_0$ is the unique solution to system (9) in Θ_Σ. The remaining parameters of model (3) can then be uniquely pinned down by the demand inverse from Lemma 1 and model (3) is globally identified.

To prove the second point, we first note that

$$\times_{t\in\mathbf{T}\setminus\mathbf{T}_0}\mathbb{R}^{J_t} \setminus \Delta^{\mathrm{ID}} = \cup_{r=1}^{R}\left[\times_{t\in\mathbf{T}\setminus\mathbf{T}_0}\mathbb{R}^{J_t} \setminus \Delta_r^{\mathrm{ID}}\right].$$

It is then sufficient to show that the Lebesgue measure of $\times_{t\in\mathbf{T}\setminus\mathbf{T}_0}\mathbb{R}^{J_t} \setminus \Delta_r^{\mathrm{ID}}$ is zero. Note that

$$\begin{aligned}
\times_{t\in\mathbf{T}\setminus\mathbf{T}_0}\mathbb{R}^{J_t} \setminus \Delta_r^{\mathrm{ID}} &= \{(\delta_{t\mathbf{J}_t})_{t\in\mathbf{T}\setminus\mathbf{T}_0} : \text{ for any } t \in \mathbf{T}\setminus\mathbf{T}_0,\ M_t s_t^{-1}(s_t(\delta_t(\Gamma^0); \Sigma_0); \Sigma_r) = \Gamma^r\} \\
&= \times_{t\in\mathbf{T}\setminus\mathbf{T}_0}\{\delta_{t\mathbf{J}_t} : M_t s_t^{-1}(s_t(\delta_t(\Gamma^0); \Sigma_0); \Sigma_r) = \Gamma^r\} \\
&= \times_{t\in\mathbf{T}\setminus\mathbf{T}_0} Z_t^r,
\end{aligned}$$

where Z_t^r is the zero set of function $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0); \Sigma_0); \Sigma_r) - \Gamma^r$. Because $\Delta_r^{\mathrm{ID}} \neq \emptyset$, there exists some $t \in \mathbf{T}\setminus\mathbf{T}_0$, for which the zero set $Z_t^r \subsetneq \mathbb{R}^{J_t}$, i.e. $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0); \Sigma_0); \Sigma_r) - \Gamma^r$ is not equal to zero for some $\delta_{t\mathbf{J}_t} \in \mathbb{R}^{J_t}$. It is then enough to show that, for this specific $Z_t^r \subsetneq \mathbb{R}^{J_t}$, the Lebesgue measure is zero.

For any Γ and Σ, because $s_t(\delta_t(\Gamma); \Sigma)$ is a composition of two real analytic functions, $\delta_t(\Gamma) : \mathbb{R}^{J_t} \to \mathbb{R}^{C_{t1}}$ and $s_t(\cdot; \Sigma) : \mathbb{R}^{C_{t1}} \to (0,1)^{C_{t1}}$ (from Theorem 3), it is itself a real analytic function from $\mathbb{R}^{J_t}$ to $(0,1)^{C_{t1}}$. Moreover, because $s_t(\cdot; \Sigma_r)$ is real analytic with respect to $\delta_t \in \mathbb{R}^{C_{t1}}$, the inverse market share function from Lemma 1, $s_t^{-1}(\cdot; \Sigma_r)$, is also real analytic with respect to $s_t' \in (0,1)^{C_{t1}}$. Then, the composition of $M_t s_t^{-1}(s_t'; \Sigma_r) - \Gamma^r$ and $s_t' = s_t(\delta_t(\Gamma^0); \Sigma_0)$ is also real analytic with respect to $\delta_{t\mathbf{J}_t} \in \mathbb{R}^{J_t}$. Consequently, Z_t^r is the zero set of the real analytic function $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0); \Sigma_0); \Sigma_r) - \Gamma^r$. There are two cases to be considered. When $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0); \Sigma_0); \Sigma_r) - \Gamma^r$ is a constant different from zero, $Z_t^r = \emptyset$ and it is of zero Lebesgue measure. Similarly, also when $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0); \Sigma_0); \Sigma_r) - \Gamma^r$ is not a constant, according to Mityagin (2015), Z_t^r is of zero Lebesgue measure.[45] This completes the proof.

7.11 Price-Setting Models Consistent with Assumption 4

In this Appendix, we show that Assumption 4 is consistent with several commonly employed pure components pricing models with any profile of demand synergies among differentiated

[45] More generally, the zero set of a non-constant analytic function defined on a P-dimensional domain can be written as the union of j-dimensional sub-manifolds, with j ranging from 0 to $P - 1$. As a consequence, the zero set is of zero Lebesgue measure. For details, see the second point of Theorem 6.3.3 (Lojasiewicz's Structure Theorem for Varieties) from Krantz and Parks (2002).

products (substitutability and/or complementarity).

Denote by $\mathbf{J}_f$ the collection of products owned by firm f and by $\mathbf{J}_{-f}$ the set of products owned by the other firms, where $\mathbf{J} = \mathbf{J}_f \cup \mathbf{J}_{-f} = 1, ..., J$ is the collection of all products available in the market. Let c_j denote the constant marginal cost of product $j \in \mathbf{J}$, $p_f = (p_j)_{j \in \mathbf{J}_f}$ the vector of prices chosen by firm f for the products it owns, and $p_{-f} = (p_k)_{k \in \mathbf{J}_{-f}}$ the vector of prices chosen by the other firms. With pure components pricing, the price of a bundle $\mathbf{b}$ is given by the sum of the prices of its individual components $p_{\mathbf{b}} = \sum_{j \in \mathbf{b}} p_j$, where each p_j is chosen by the firm that owns it. Then, under pure components pricing, the profit function of firm f takes the following form:

$$\pi_f(p_f, p_{-f}) = \sum_{j \in f} s_{j.}(p_{\mathbf{J}})(p_j - c_j), \tag{53}$$

where $s_{j.}(p_{\mathbf{J}}) = \sum_{\mathbf{b}:\mathbf{b} \ni j} s_{\mathbf{b}}(p_{\mathbf{J}})$ is the marginal market share function of product j and $p_{\mathbf{J}} = (p_1, ..., p_J)$. Further denote the ownership matrix $\Omega = (a_{jj'})_{j,j'=1,...,J}$ where $a_{jj'} = 1$ if j and j' are owned by the same firm and 0 otherwise. Under complete information, the necessary first-order conditions for a Bertrand-Nash equilibrium in pure components are:

$$\left[\frac{\partial s_{\mathbf{J}.}}{\partial p_{\mathbf{J}}} \otimes \Omega\right](p_{\mathbf{J}} - c_{\mathbf{J}}) + s_{\mathbf{J}.}(p_{\mathbf{J}}) = 0, \tag{54}$$

where $\otimes$ is the element-wise multiplication (Kronecker product), $s_{\mathbf{J}.} = (s_{j.}(p_{\mathbf{J}}))_{j \in \mathbf{J}}$ is the vector of marginal market share functions, $p_{\mathbf{J}} = (p_j)_{j \in \mathbf{J}}$, and $c_{\mathbf{J}} = (c_j)_{j \in \mathbf{J}}$. For different configurations of the ownership matrix, (54) can be specialized to different market structures such as monopoly, duopoly, and oligopoly.

The identifiability of $c_{\mathbf{J}}$ is determined by the invertibility of the matrix $\frac{\partial s_{\mathbf{J}.}}{\partial p_{\mathbf{J}}} \otimes \Omega$. As long as $\frac{\partial s_{\mathbf{J}.}}{\partial p_{\mathbf{J}}} \otimes \Omega$ is invertible, we obtain:

$$c_{\mathbf{J}} = p_{\mathbf{J}} + \left[\frac{\partial s_{\mathbf{J}.}}{\partial p_{\mathbf{J}}} \otimes \Omega\right]^{-1} s_{\mathbf{J}.}(p_{\mathbf{J}}).$$

We now show that for any ownership matrix, $\frac{\partial s_{\mathbf{J}.}}{\partial p_{\mathbf{J}}} \otimes \Omega$ is invertible. Let $p = (p_{\mathbf{J}}, (p_{\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_2})$ denote the vector of prices for all bundles in the choice set. Moreover, we assume that p_j enters linearly in $u_{ij} = \delta_j + \mu_{ij}(\beta_i)$ with individual-specific coefficient $\alpha_i < 0$, which is part of the vector of random coefficients β_i. Then, by using the notation M introduced in the proof of Theorem 4, we can write:

$$\begin{aligned}\frac{\partial s_{\mathbf{J}.}}{\partial p_{\mathbf{J}}} &= \int M \frac{\partial s_i(\beta_i)}{\partial p_{\mathbf{J}}} dF(\beta_i) \\ &= \int \alpha_i M \frac{\partial s_i(\beta_i)}{\partial u_i} M^{\mathrm{T}} dF(\beta_i),\end{aligned} \tag{55}$$

where $u_i = (\delta_{\mathbf{b}} + \mu_{i\mathbf{b}}(\beta_i))_{\mathbf{b}\in\mathbf{C}_1}$. As shown in the proof of Lemma 1 (see Appendix 7.3), $\frac{\partial s_i(\beta_i)}{\partial u_i}$ is positive-definite for any β_i. Moreover, M is full row rank and therefore M^{T} is of full column rank. Consequently, $M\frac{\partial s_i(\beta_i)}{\partial u_i}M^{\mathrm{T}}$ is positive-definite for any β_i. Because $\alpha_i < 0$, $\frac{\partial s_{\mathbf{J}.}}{\partial p_{\mathbf{J}}}$ is negative-definite. Note that Ω is a symmetric block diagonal matrix that contains only 1s and 0s. Then, $\frac{\partial s_{\mathbf{J}.}}{\partial p_{\mathbf{J}}} \otimes \Omega$ is also block diagonal. Because each block is a principal submatrix of $\frac{\partial s_{\mathbf{J}.}}{\partial p_{\mathbf{J}}}$, these blocks are also negative-definite. Then, $\frac{\partial s_{\mathbf{J}.}}{\partial p_{\mathbf{J}}} \otimes \Omega$ is negative-definite and thus invertible.

7.12 Proof of Theorem 5

To ease exposition, we use the same notation for M_t introduced in the proof of Theorem 4. The proof of the first point is similar to that of Theorem 4. We first use Lemma 1 and re-write Ξ_r^{ID} as:

$$\begin{aligned}\Xi_r^{\mathrm{ID}} = \{(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})_{t\in\mathbf{T}\setminus\mathbf{T}_0} :& \exists t \in \mathbf{T}\setminus\mathbf{T}_0 \text{ such that } M_t s_t^{-1}(s_t(\delta_t(\Gamma^0); p_{t\mathbf{J}_t}, \Sigma_0); p_{t\mathbf{J}_t}, \Sigma_r) \neq \Gamma^r \\ & \text{for any } p_{t\mathbf{J}_t} \in p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})\}.\end{aligned}$$

On the one hand, when $(\xi_{t\mathbf{J}_t}^0, c_{t\mathbf{J}_t}^0)_{t\in\mathbf{T}\setminus\mathbf{T}_0} \in \Xi^{\mathrm{ID}} = \cap_{r=1}^R \Xi_r^{\mathrm{ID}}$, for any $r = 1,..,R$ there exists some market $t^+ \in \mathbf{T}\setminus\mathbf{T}_0$ such that $M_{t^+}s_{t^+}^{-1}(\mathcal{s}_{t^+}; p_{t\mathbf{J}_t}, \Sigma_r) \neq \Gamma^r$. On the other, according to the definition of Σ_r, $m(\Sigma_r; \mathbf{T}_0) = 0$. Then, by construction $m(\Sigma_r; \mathbf{T}_0 \cup \{t^+\}) \neq 0$ and therefore $m(\Sigma_r; \mathbf{T}) \neq 0$ for $r = 1, ..., R$.

Remember that the set of solutions to $m(\Sigma'; \mathbf{T}_0) = 0$ in Θ_Σ is $S(\Sigma) = \{\Sigma_r : r = 0, ..., R\}$. Consequently, the set of solutions to $m(\Sigma'; \mathbf{T}) = 0$ is a subset of $S(\Sigma)$. Given that $m(\Sigma_r; \mathbf{T}) \neq 0$ for $r = 1, ..., R$, and that $m(\Sigma_0; \mathbf{T}) = 0$, $\Sigma' = \Sigma_0$ is the unique solution to system (9) in Θ_Σ. The remaining parameters of model (3) can then be uniquely pinned down by the demand inverse from Lemma 1 and model (3) is globally identified.

To prove the second point, we first note that

$$\times_{t\in\mathbf{T}\setminus\mathbf{T}_0}[D_{t\xi} \times D_{tc}] \setminus \Xi^{\mathrm{ID}} = \cup_{r=1}^R \left[\times_{t\in\mathbf{T}\setminus\mathbf{T}_0}[D_{t\xi} \times D_{tc}] \setminus \Xi_r^{\mathrm{ID}}\right].$$

It is then sufficient to show that the Lebesgue measure of $\times_{t\in\mathbf{T}\setminus\mathbf{T}_0}[D_{t\xi} \times D_{tc}] \setminus \Xi_r^{\mathrm{ID}}$ is zero. Note that

$$\begin{aligned}\times_{t\in\mathbf{T}\setminus\mathbf{T}_0}[D_{t\xi} \times D_{tc}] \setminus \Xi_r^{\mathrm{ID}} &= \{(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})_{t\in\mathbf{T}\setminus\mathbf{T}_0} : \\ &\qquad \text{for some } t \in \mathbf{T}\setminus\mathbf{T}_0,\ M_t s_t^{-1}(s_t(\delta_t(\Gamma^0); p_{t\mathbf{J}_t}, \Sigma_0); p_{t\mathbf{J}_t}, \Sigma_r) = \Gamma^r \\ &\qquad \text{for some } p_{t\mathbf{J}_t} \in p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})\} \\ &= \times_{t\in\mathbf{T}\setminus\mathbf{T}_0}\{(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}) :\ \Gamma^r \in M_t s_t^{-1}(s_t(\delta_t(\Gamma^0); p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}), \Sigma_0); p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}), \Sigma_r) \\ &= \times_{t\in\mathbf{T}\setminus\mathbf{T}_0} Z_t^{+r},\end{aligned}$$

where Z_t^{+r} is the zero set of $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$ such that $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0); p_{t\mathbf{J}_t}, \Sigma_0); p_{t\mathbf{J}_t}, \Sigma_r) - \Gamma^r = 0$ for

some $p_{t\mathbf{J}_t} \in p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$. Note that Z_t^{+r} is measurable for $t \in \mathbf{T}$ and $r = 1, ..., R$. Given that for any $r = 1, ..., R$, there exists $t \in \mathbf{T} \setminus \mathbf{T}_0$, for any $c_{t\mathbf{J}_t} \in D_{tc}$, such that $\Xi_{tr}^{\mathrm{ID}}(c_{t\mathbf{J}_t}) \neq \emptyset$; we will prove that, for this specific market t, Z_t^{+r} is of zero Lebesgue measure.

Note that the Lebesgue measure of Z_t^{+r} in $D_{t\xi} \times D_{tc}$ is

$$\begin{aligned} m(Z_t^{+r}) &= \int_{D_{t\xi} \times D_{tc}} \mathbf{1}\{Z_t^{+r}\} d(c_{t\mathbf{J}_t}, \xi_{t\mathbf{J}_t}) \\ &= \int_{D_{t\xi} \times D_{tc}} \mathbf{1}\{(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}) :\ \Gamma^r \in M_t s_t^{-1}(s_t(\delta_t(\Gamma^0); p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}), \Sigma_0); p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}), \Sigma_r)\} d(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}), \end{aligned}$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function. Define $\Phi : (\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t}) \to (\xi_{t\mathbf{J}_t}, \phi(\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t}))$. According to Assumption 4, Φ is a C^1 mapping from $(\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t}) \in \{(\xi'_{t\mathbf{J}_t}, p'_{t\mathbf{J}_t}) : \xi'_{t\mathbf{J}_t} \in D_{t\xi}, p'_{t\mathbf{J}_t} \in \mathbf{P}_t(\xi_{t\mathbf{J}_t})\}$ to $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}) \in D_{t\xi} \times D_{tc}$ and onto. Let $\mathrm{Card}[\Phi^{-1}](\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$ denote the cardinality of the inverse image of Φ at $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$. Note that $\mathrm{Card}[\Phi^{-1}](\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$ is equal to the number of Nash equilibria of the pricing game at $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$ and therefore belongs to $\in \mathbb{N}_+ \cup \{\infty\}$ according to Assumption 4. Then, by using Theorem 1.16-2 of Ciarlet (2013) and Fubini's Theorem, we obtain:

$$\begin{aligned} m(Z_t^{+r}) &\leq \int_{D_{t\xi} \times D_{tc}} \mathbf{1}\{(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}) :\ \Gamma^r \in M_t s_t^{-1}(s_t(\delta_t(\Gamma^0); p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}), \Sigma_0); p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}), \Sigma_r)\} \mathrm{Card}[\Phi^{-1}](\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}) d(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}) \\ &= \int_{\{(\xi'_{t\mathbf{J}_t}, p'_{t\mathbf{J}_t}) : \xi'_{t\mathbf{J}_t} \in D_{t\xi}, p'_{t\mathbf{J}_t} \in \mathbf{P}_t(\xi_{t\mathbf{J}_t})\}} \mathbf{1}\{(\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t}) :\ \Gamma^r = M_t s_t^{-1}(s_t(\delta_t(\Gamma^0); p_{t\mathbf{J}_t}, \Sigma_0); p_{t\mathbf{J}_t}, \Sigma_r)\} \left| \frac{\partial \Phi_{t\mathbf{J}_t}}{\partial(\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t})}(\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t}) \right| d(\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t}) \\ &= \int_{\mathbf{P}_t} \left[\int_{\mathbf{P}_t(\xi_{t\mathbf{J}_t}) \ni p_{t\mathbf{J}_t}} \mathbf{1}\{\xi_{t\mathbf{J}_t} :\ \Gamma^r = M_t s_t^{-1}(s_t(\delta_t(\Gamma^0); p_{t\mathbf{J}_t}, \Sigma_0); p_{t\mathbf{J}_t}, \Sigma_r)\} \left| \frac{\partial \phi_{t\mathbf{J}_t}}{\partial p_{t\mathbf{J}_t}}(\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t}) \right| d\xi_{t\mathbf{J}_t} \right] dp_{t\mathbf{J}_t}. \end{aligned}$$

Because $\delta_{t\mathbf{J}_t}(\Delta\delta_{t\mathbf{J}_t}, \xi_{t\mathbf{J}_t}) = \Delta\delta_{t\mathbf{J}_t} + \xi_{t\mathbf{J}_t}$, given $p_{t\mathbf{J}_t}$ (and therefore $\Delta\delta_{t\mathbf{J}_t}$) and by applying Theorem 3, we obtain that the market share function $s_t(\delta_t(\Gamma'); p_{t\mathbf{J}_t}, \Sigma')$ is also real analytic with respect to $\xi_{t\mathbf{J}_t} \in \mathbb{R}^{J_t}$. Moreover, for any $p_{t\mathbf{J}_t} \in \mathbf{P}_t$, there exists a $c_{t\mathbf{J}_t}$ and $\xi_{t\mathbf{J}_t}$ such that $p_{t\mathbf{J}_t} \in p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$. Therefore $\Xi_{tr}(c_{t\mathbf{J}_t}) \neq \emptyset$ implies that $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0); p_{t\mathbf{J}_t}, \Sigma_0); p_{t\mathbf{J}_t}, \Sigma_r) \neq \Gamma^r$ for some $\xi_{t\mathbf{J}_t} \in D_{t\xi}$, i.e., $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0); p_{t\mathbf{J}_t}, \Sigma_0); p_{t\mathbf{J}_t}, \Sigma_r) - \Gamma^r$ is not constantly zero in $D_{t\xi}$. Similarly to the proof of the second statement of Theorem 4, $\{\xi_{t\mathbf{J}_t} :\ M_t s_t^{-1}(s_t(\delta_t(\Gamma^0); p_{t\mathbf{J}_t}, \Sigma_0); p_{t\mathbf{J}_t}, \Sigma_r) = \Gamma^r\}$ has thus zero Lebesgue measure in $D_{t\xi}$ and

$$\mathbf{1}\{\xi_{t\mathbf{J}_t} :\ M_t s_t^{-1}(s_t(\delta_t(\Gamma^0); p_{t\mathbf{J}_t}, \Sigma_0); p_{t\mathbf{J}_t}, \Sigma_r) = \Gamma^r\} \left| \frac{\partial \phi_{t\mathbf{J}_t}}{\partial p_{t\mathbf{J}_t}}(p_{t\mathbf{J}_t}; \xi_{t\mathbf{J}_t}) \right| = 0 \text{ almost everywhere.}$$

Consequently,

$$\int_{\mathbf{P}_t(\xi_{t\mathbf{J}_t}) \ni p_{t\mathbf{J}_t}} \mathbf{1}\{\xi_{t\mathbf{J}_t} :\ M_t s_t^{-1}(s_t(\delta_t(\Gamma^0); p_{t\mathbf{J}_t}, \Sigma_0); p_{t\mathbf{J}_t}, \Sigma_r) = \Gamma^r\} \left| \frac{\partial c_{t\mathbf{J}_t}}{\partial p_{t\mathbf{J}_t}}(p_{t\mathbf{J}_t}; \xi_{t\mathbf{J}_t}, \Gamma^0, \Sigma_0) \right| d\xi_{t\mathbf{J}_t} = 0,$$

and finally $m(Z_t^{+r}) \leq 0$. Consequently, $m(Z_t^{+r}) = 0$ and this completes the proof.

7.13 Proof of Lemma 3

We first show that $\ell(\delta'_{1\mathbf{J}_1},...,\delta'_{I\mathbf{J}_T},\Gamma',\Sigma')$ is maximized at the true parameters $(\delta_{1\mathbf{J}_1},...,\delta_{I\mathbf{J}_T},\Gamma,\Sigma)$. Note that for any $t=1,...,T$ and $\mathbf{b}\in\mathbf{C}_t$, $\mathcal{s}_{t\mathbf{b}}=s_{t\mathbf{b}}(\delta_t(\Gamma);F(\cdot;\Sigma))$. Then, by using Jensen's inequality, we have for any $(\delta'_{1\mathbf{J}_1},...,\delta'_{I\mathbf{J}_T},\Gamma',\Sigma')$,

$$\begin{aligned}\ell(\delta'_{1\mathbf{J}_1},...,\delta'_{T\mathbf{J}_T},\Gamma',\Sigma')-\ell(\delta_{1\mathbf{J}_1},...,\delta_{T\mathbf{J}_T},\Gamma,\Sigma)&=\sum_{t=1}^{T}\sum_{\mathbf{b}\in\mathbf{C}_t}\mathcal{s}_{t\mathbf{b}}\log\frac{s_{t\mathbf{b}}(\delta'_t(\Gamma');F(\cdot;\Sigma'))}{s_{t\mathbf{b}}(\delta_t(\Gamma);F(\cdot;\Sigma))}\\&\le\sum_{t=1}^{T}\log\sum_{\mathbf{b}\in\mathbf{C}_t}\mathcal{s}_{t\mathbf{b}}\frac{s_{t\mathbf{b}}(\delta'_t(\Gamma');F(\cdot;\Sigma'))}{s_{t\mathbf{b}}(\delta_t(\Gamma);F(\cdot;\Sigma))}\\&\le 0.\end{aligned}\tag{56}$$

We now show the uniqueness by contradiction. Suppose that there exists a $(\tilde{\delta}_{1\mathbf{J}_1},...,\tilde{\delta}_{t\mathbf{J}_T},\tilde{\Gamma},\tilde{\Sigma})\neq(\delta_{1\mathbf{J}_1},...,\delta_{T\mathbf{J}_T},\Gamma,\Sigma)$ such that $(\tilde{\delta}_{1\mathbf{J}_1},...,\tilde{\delta}_{T\mathbf{J}_T},\tilde{\Gamma},\tilde{\Sigma})$ is also a maximizer of $\ell(\delta',\Gamma',\Sigma')$. According to Jensen's inequality (56), this is equivalent to having $s_{t\mathbf{b}}(\tilde{\delta}_t(\tilde{\Gamma});\tilde{\Sigma})=\mathcal{s}_{t\mathbf{b}}$ for each $t=1,...,T$ and $\mathbf{b}\in\mathbf{C}_{t1}$. As a consequence, we have $m_{\mathbf{b}}(\tilde{\Sigma};\mathbf{T})=0$ and hence $m(\tilde{\Sigma};\mathbf{T})=0$ in addition to $m(\Sigma;\mathbf{T})=0$. Note that $\tilde{\Sigma}\neq\Sigma$. Otherwise, by Lemma 1, $\tilde{\delta}_{t\mathbf{J}_t}=\delta_{t\mathbf{J}_t}$ and $\tilde{\Gamma}=\Gamma$ and this would be inconsistent with $(\tilde{\delta}_{1\mathbf{J}_1},...,\tilde{\delta}_{t\mathbf{J}_T},\tilde{\Gamma},\tilde{\Sigma})\neq(\delta_{1\mathbf{J}_1},...,\delta_{T\mathbf{J}_T},\Gamma,\Sigma)$. However, because the true $(\delta_{t\mathbf{J}_t})_{t\in\mathbf{T}\setminus\mathbf{T}_0}\in\Delta^{\mathrm{ID}}$, Theorem 4 rules out the possibility of having any other $\tilde{\Sigma}$ different from Σ for which system (9) holds, giving rise to a contradiction. This completes the proof.

7.14 Proof of Theorem 6

We assume the following regularity conditions.

1. θ is an interior point of Θ;
2. the market share function $s_{t\mathbf{b}}$, $t=1,...,T$ and $\mathbf{b}\in\mathbf{C}_t$, is twice continuously differentiable with respect to θ' in Θ;
3. $\sqrt{I}(\hat{\mathcal{s}}_t-\mathcal{s}_t)\xrightarrow{d}\mathcal{N}(0,\Omega_t)$ independently for $t=1,...,T$, where Ω_t is positive-definite;
4. $\sum_{t=1}^{T}G_t\Omega_tG_t^{\mathrm{T}}$ is positive-definite, where $G_t=\left(\left[\frac{\partial\log s_{t\mathbf{b}}}{\partial\theta'}-\frac{\partial\log s_{t0}}{\partial\theta'}\right]\Big|_{\theta'=\theta}\right)_{\mathbf{b}\in\mathbf{C}_{t1}}$.
5. $\frac{\partial^2\ell(\theta')}{\partial\theta'^2}\Big|_{\theta'=\theta}$ is non-singular.

Note that Condition 3 encompasses the case where individuals in market t make independent purchase decisions. Condition 4 can be obtained when G_t is a full row rank matrix for each $t=1,...,T$. Our proof for the consistency statement is mainly based on Theorem 2.1 by Newey and McFadden (1994), according to which we need to verify four conditions.

1. $\theta = (\delta_{1\mathbf{J}_1}, ..., \delta_{I\mathbf{J}_I}, \Gamma, \Sigma)$ **is the unique maximizer of** $\ell(\theta')$ **in** Θ**.** Under Assumptions 1–3, this is guaranteed by Lemma 3.

2. Θ **is compact.** This is guaranteed by the definition of Θ.

3. $\ell(\theta')$ **is continuous with respect to** θ' **in** Θ**.** Because $F(\cdot;\Sigma')$ is a continuous distribution with a continuous density function with respect to Σ', and individual i's choice probability $s_{i\mathbf{b}}$ is a continuous function of δ'_t and Γ', according to the definition of market share function $s_{t\mathbf{b}}(\delta'_t(\Gamma'); F(\cdot;\Sigma'))$ in (3), $s_{t\mathbf{b}}(\delta'_t(\Gamma'); F(\cdot;\Sigma'))$ is continuous with respect to θ'. Then, $\ell(\theta')$ is continuous with respect to θ'.

4. $\sup_{\theta'\in\Theta} |\ell_I(\theta'; \hat{\mathscr{s}}_1, ..., \hat{\mathscr{s}}_T) - \ell(\theta')| \xrightarrow{p} 0$**.** To see this, note that:

$$\begin{aligned}
&\sup_{\theta'\in\Theta} |\ell_I(\theta'; \hat{\mathscr{s}}_1, ..., \hat{\mathscr{s}}_T) - \ell(\theta')| \\
&= \sup_{\theta'\in\Theta} \left| \sum_{t=1}^{T} \sum_{\mathbf{b}\in\mathbf{C}_t} \hat{\mathscr{s}}_{t\mathbf{b}} \log s_{t\mathbf{b}}(\delta'_t(\Gamma'); F(\cdot;\Sigma')) - \sum_{t=1}^{T} \sum_{\mathbf{b}\in\mathbf{C}_t} \mathscr{s}_{t\mathbf{b}} \log s_{t\mathbf{b}}(\delta'_t(\Gamma'); F(\cdot;\Sigma')) \right| \\
&\leq \sup_{\substack{\theta'\in\Theta \\ t=1,...,T,\mathbf{b}\in\mathbf{C}_t}} |\log s_{t\mathbf{b}}(\delta'_t(\Gamma'); F(\cdot;\Sigma'))| \sum_{t=1}^{T} \sum_{\mathbf{b}\in\mathbf{C}_t} |\hat{\mathscr{s}}_{t\mathbf{b}} - \mathscr{s}_{t\mathbf{b}}|.
\end{aligned} \tag{57}$$

Because $\log s_{t\mathbf{b}}(\delta'_t(\Gamma'); F(\cdot;\Sigma'))$ is continuous in Θ and Θ is compact, $\log s_{t\mathbf{b}}(\delta'_t(\Gamma'); F(\cdot;\Sigma'))$ is uniformly bounded in Θ. Moreover, because both the number of markets, T, and the number of bundles in each market are finite, $\sup_{\substack{\theta'\in\Theta \\ t=1,...,T,\mathbf{b}\in\mathbf{C}_t}} |\log s_{t\mathbf{b}}(\delta'_t(\Gamma'); F(\cdot;\Sigma'))| < \infty$. Note that $\hat{\mathscr{s}}_{t\mathbf{b}} \xrightarrow{p} \mathscr{s}_{t\mathbf{b}}$ for $t = 1, ..., T$ and $\mathbf{b} \in \mathbf{C}_t$. Then, the right-hand side of (57) converges to zero in probability. As a consequence, $\sup_{\theta'\in\Theta} |\ell_I(\theta'; \hat{\mathscr{s}}_1, ..., \hat{\mathscr{s}}_T) - \ell(\theta')|$ converges to zero in probability.

According to Theorem 2.1 by Newey and McFadden (1994), the four conditions verified above guarantee the consistency of the MLE.

The proof of asymptotic normality is mainly based on Theorem 3.1 by Newey and McFadden (1994), according to which we need to verify the following six conditions.

1. $\hat{\theta} \xrightarrow{p} \theta$**.** This has just been shown above.

2. θ **is an interior point of** Θ**.** This is guaranteed by regularity condition 1.

3. $\ell_I(\theta'; \hat{\mathscr{s}}_1, ..., \hat{\mathscr{s}}_T)$ **is twice continuously differentiable in** Θ**.** According to regularity condition 2, the market share function $s_{t\mathbf{b}}$, $t = 1, ..., T$ and $\mathbf{b} \in \mathbf{C}_t$, $s_{t\mathbf{b}}$ is twice continuously differentiable with respect to θ'. Then, $\ell_I(\theta'; \hat{\mathscr{s}}_1, ..., \hat{\mathscr{s}}_T)$ is twice continuously differentiable in Θ.

4. $\sqrt{I}\frac{\partial \ell_I}{\partial \theta'}\big|_{\theta'=\theta}$ **converges to a centered normal distribution.** Define $\ell^t(\theta') = \sum_{\mathbf{b}\in\mathbf{C}_t} \jmath_{t\mathbf{b}} \log s_{t\mathbf{b}}(\theta')$. Note that $\ell^t(\theta')$ is maximized at $\theta' = \theta$, for $t = 1, ..., T$. As a consequence, $\frac{\partial \ell^t}{\partial \theta'}\big|_{\theta'=\theta} = 0$ for $t = 1, ..., T$. Then,

$$
\begin{aligned}
\sqrt{I}\frac{\partial \ell_I}{\partial \theta'}\bigg|_{\theta'=\theta} &= \sqrt{I}\sum_{t=1}^{T}\sum_{\mathbf{b}\in\mathbf{C}_t} \hat{\jmath}_{t\mathbf{b}} \frac{\log \partial s_{t\mathbf{b}}}{\partial \theta'}\bigg|_{\theta'=\theta} \\
&= \sqrt{I}\sum_{t=1}^{T}\left[\sum_{\mathbf{b}\in\mathbf{C}_t} \hat{\jmath}_{t\mathbf{b}} \frac{\partial \log s_{t\mathbf{b}}}{\partial \theta'}\bigg|_{\theta'=\theta} - \frac{\partial \ell^t}{\partial \theta'}\bigg|_{\theta'=\theta}\right] \\
&= \sqrt{I}\sum_{t=1}^{T}\left[\sum_{\mathbf{b}\in\mathbf{C}_{t1}} [\hat{\jmath}_{t\mathbf{b}} - \jmath_{t\mathbf{b}}] \frac{\partial \log s_{t\mathbf{b}}}{\partial \theta'}\bigg|_{\theta'=\theta} + [\hat{\jmath}_{t0} - \jmath_{t0}] \frac{\partial \log s_{t0}}{\partial \theta'}\bigg|_{\theta'=\theta}\right] \\
&= \sum_{t=1}^{T}\sum_{\mathbf{b}\in\mathbf{C}_t} \sqrt{I}\,[\hat{\jmath}_{t\mathbf{b}} - \jmath_{t\mathbf{b}}]\left[\frac{\partial \log s_{t\mathbf{b}}}{\partial \theta'} - \frac{\partial \log s_{t0}}{\partial \theta'}\right]\bigg|_{\theta'=\theta} \\
&= \sum_{t=1}^{T}\left(\left[\frac{\partial \log s_{t\mathbf{b}}}{\partial \theta'} - \frac{\partial \log s_{t0}}{\partial \theta'}\right]\bigg|_{\theta'=\theta}\right)_{\mathbf{b}\in\mathbf{C}_{t1}} \sqrt{I}\,[\hat{\jmath}_t - \jmath_t]\,.
\end{aligned} \tag{58}
$$

According to regularity condition 3, $\sqrt{I}\,[\hat{\jmath}_t - \jmath_t] \xrightarrow{d} \mathcal{N}(0, \Omega_t)$ independently for $t = 1, ..., T$, by using Slutsky's Theorem, one obtains

$$
\sum_{t=1}^{T}\left(\left[\frac{\partial \log s_{t\mathbf{b}}}{\partial \theta'} - \frac{\partial \log s_{t0}}{\partial \theta'}\right]\bigg|_{\theta'=\theta}\right)_{\mathbf{b}\in\mathbf{C}_{t1}} \sqrt{I}\,[\hat{\jmath}_t - \jmath_t] \xrightarrow{d} \mathcal{N}(0, \sum_{t=1}^{T} G_t \Omega_t G_t^{\mathrm{T}}), \tag{59}
$$

where $G_t = \left(\left[\frac{\partial \log s_{t\mathbf{b}}}{\partial \theta'} - \frac{\partial \log s_{t0}}{\partial \theta'}\right]\big|_{\theta'=\theta}\right)_{\mathbf{b}\in\mathbf{C}_{t1}}$. $\sum_{t=1}^{I} G_t \Omega_t G_t^{\mathrm{T}}$ is positive-definite according to regularity condition 4.

5. $\sup_{\theta'\in\Theta}\left|\frac{\partial^2 \ell_I}{\partial \theta'^2}(\theta') - H(\theta')\right| \xrightarrow{p} 0$**.** Define $H(\theta') = \sum_{t=1,...,T}\sum_{\mathbf{b}\in\mathbf{C}_t} \jmath_{t\mathbf{b}} \frac{\partial \log^2 s_{t\mathbf{b}}(\theta')}{\partial \theta'^2}$. Then, under regularity condition 2, $H(\theta')$ is continuous in Θ. Note that, similarly to (57), we have:

$$
\sup_{\theta'\in\Theta}\left|\frac{\partial^2 \ell_I}{\partial \theta'^2}(\theta') - H(\theta')\right| \le \sup_{\theta'\in\Theta}\left|\frac{\partial \log^2 s_{t\mathbf{b}}(\theta')}{\partial \theta'^2}\right| \sum_{t=1,...,T,\mathbf{b}\in\mathbf{C}_t} |\hat{\jmath}_{t\mathbf{b}} - \jmath_{t\mathbf{b}}|. \tag{60}
$$

Due to the continuity of $\frac{\partial \log^2 s_{t\mathbf{b}}(\theta')}{\partial \theta'^2}$ in the compact set Θ, for $t = 1, ..., T$, $\sup_{\theta'\in\Theta}\left|\frac{\partial \log^2 s_{t\mathbf{b}}(\theta')}{\partial \theta'^2}\right| < \infty$. Because $\hat{\jmath}_t \xrightarrow{p} \jmath_t$ for $t = 1, ..., T$, the right-hand side of (60) converges to zero in probability and $\sup_{\theta'\in\Theta}\left|\frac{\partial^2 \ell_I}{\partial \theta'^2}(\theta') - H(\theta')\right| \xrightarrow{p} 0$.

6. $H(\theta) = \frac{\partial^2 \ell(\theta')}{\partial \theta'^2}\big|_{\theta'=\theta}$ **is non-singular.** This is ensured by regularity condition 5.
All the six conditions of Theorem 3.1 by Newey and McFadden (1994) are satisfied and there

exists $W = H(\theta)^{-1}[\sum_{t=1}^{T} G_t \Omega_t G_t^{\mathrm{T}}] H(\theta)^{-1}$ such that $\sqrt{I}(\hat{\theta} - \theta) \xrightarrow{d} \mathscr{N}(0, W)$. This completes the proof.

7.15 Proof of Theorem 7

To ease exposition, we use the same notation for M_t introduced in the proof of Theorem 4. Note that M_t is of full row rank and therefore M_t^{T} is of full column rank. Without loss of generality, we prove Theorem 7 for market t and for any $\Sigma' \in \Theta_\Sigma$ and $\Gamma' \in \cap_{t=1}^{T} \bar{\Theta}_\Gamma^t(\Sigma')$.

Define $\mathbf{S}_{t2}(\mathscr{s}_{t\mathbf{J}_t \cdot}) = \{\mathscr{s}'_{t\mathbf{C}_{t2}} : \mathscr{s}'_{t\mathbf{C}_{t2}} = (\mathscr{s}'_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_{t2}}, \mathscr{s}'_{t\mathbf{b}} > 0, -M_t^{1\mathrm{T}} \mathscr{s}'_{t\mathbf{C}_{t2}} < \mathscr{s}_{t\mathbf{J}_t \cdot}, (M_t^{1\mathrm{T}} \mathscr{s}'_{t\mathbf{C}_{t2}} + \mathscr{s}_{t\mathbf{J}_t \cdot}, \mathscr{s}'_{t\mathbf{C}_{t2}})^{\mathrm{T}} \mathbf{1} < 1\}$ as the collection of admissible vectors of market shares of multi-product bundles consistent with the observed marginals $\mathscr{s}_{t\mathbf{J}_t \cdot}$. Given any $\mathscr{s}'_{t\mathbf{C}_{t2}} \in \mathbf{S}_{t2}(\mathscr{s}_{t\mathbf{J}_t \cdot})$ and observed marginals $\mathscr{s}_{t\mathbf{J}_t \cdot}$, we can construct an admissible vector of market shares of all bundles $\mathscr{s}'_t = ((\mathscr{s}'_{tj})_{j \in \mathbf{J}_t}, \mathscr{s}'_{t\mathbf{C}_{t2}})$, where $\mathscr{s}'_{tj} = \mathscr{s}_{tj\cdot} - \sum_{\mathbf{b} \in \mathbf{C}_{t2} : j \in \mathbf{b}} \mathscr{s}'_{t\mathbf{b}}$. Because of Lemma 1, given Σ' we can invert $\mathscr{s}'_t$ and obtain the corresponding $\delta'_t \in \mathbb{R}^{C_{t1}}$ as a function of:

$$
\begin{aligned}
\delta'_t &= ((\delta'_{tj})_{j \in \mathbf{J}_t}, (\delta'_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_{t2}})^{\mathrm{T}} \\
&= s_t^{-1}(\mathscr{s}'_t; \Sigma') \\
&= s_t^{-1}((\mathscr{s}_{tj\cdot} - \sum_{\mathbf{b} \in \mathbf{C}_{t2} : j \in \mathbf{b}} \mathscr{s}'_{t\mathbf{b}})_{j \in \mathbf{J}_t}, \mathscr{s}'_{t\mathbf{C}_{t2}}; \Sigma'), \text{ where} \\
\delta'_{tj} &= s_{tj}^{-1}((\mathscr{s}_{tj\cdot} - \sum_{\mathbf{b} \in \mathbf{C}_{t2} : j \in \mathbf{b}} \mathscr{s}'_{t\mathbf{b}})_{j \in \mathbf{J}_t}, \mathscr{s}'_{t\mathbf{C}_{t2}}; \Sigma'), \\
\delta'_{t\mathbf{b}} &= s_{t\mathbf{b}}^{-1}((\mathscr{s}_{tj\cdot} - \sum_{\mathbf{b} \in \mathbf{C}_{t2} : j \in \mathbf{b}} \mathscr{s}'_{t\mathbf{b}})_{j \in \mathbf{J}_t}, \mathscr{s}'_{t\mathbf{C}_{t2}}; \Sigma').
\end{aligned}
\tag{61}
$$

By using M_t, we can recover an admissible Γ'_t from δ'_t by:

$$
\begin{aligned}
\Gamma'_t &= M_t \delta'_t \\
&= M_t s_t^{-1}((\mathscr{s}_{tj\cdot} - \sum_{\mathbf{b} \in \mathbf{C}_{t2} : j \in \mathbf{b}} \mathscr{s}'_{t\mathbf{b}})_{j \in \mathbf{J}_t}, \mathscr{s}'_{t\mathbf{C}_{t2}}; \Sigma') \\
&= M_t s_t^{-1}(M_t^{\mathrm{T}} \mathscr{s}'_{t\mathbf{C}_{t2}} + (\mathscr{s}_{t\mathbf{J}_t \cdot}^{\mathrm{T}}, 0, ..., 0)^{\mathrm{T}}; \Sigma').
\end{aligned}
$$

Consequently, for any $\mathscr{s}'_{t\mathbf{C}_{t2}}$ there exists a $\Gamma'_t = \Gamma_t(\mathscr{s}'_{t\mathbf{C}_{t2}}; \mathscr{s}_{t\mathbf{J}_t \cdot}, \Sigma')$ such that (61) holds.

We now compute from (61) the derivative of $\Gamma'_t = \Gamma_t(\mathscr{s}'_{t\mathbf{C}_{t2}}; \mathscr{s}_{t\mathbf{J}_t \cdot}, \Sigma')$ with respect to $\mathscr{s}'_{t\mathbf{C}_{t2}}$:

$$
\begin{aligned}
\frac{d\Gamma_t}{d\mathscr{s}'_{t\mathbf{C}_{t2}}} &= M_t \frac{\partial s_t^{-1}}{\partial \mathscr{s}'_t}(M_t^{\mathrm{T}} \mathscr{s}'_{t\mathbf{C}_{t2}} + (\mathscr{s}_{t\mathbf{J}_t \cdot}^{\mathrm{T}}, 0, ..., 0)^{\mathrm{T}}; \Sigma') M_t^{\mathrm{T}} \\
&= M_t \left[\frac{\partial s_t}{\partial \delta'_t}(\delta'_t; \Sigma') \right]^{-1} M_t^{\mathrm{T}}.
\end{aligned}
\tag{62}
$$

Because $\frac{\partial s_t}{\partial \delta_t'}(\delta_t'; \Sigma')$ is positive-definite and M_t^{T} is of full column rank, $\frac{d\Gamma_t}{d\mathscr{s}'_{t\mathbf{C}_{t2}}}$ is also positive-definite and therefore positive quasi-definite for any $\mathscr{s}'_{t\mathbf{C}_{t2}} \in \mathbf{S}_{t2}(\mathscr{s}_{t\mathbf{J}_t\cdot})$. Note that $\mathbf{S}_{t2}(\mathscr{s}_{t\mathbf{J}_t\cdot})$ is convex. According to Gale and Nikaido (1965), $\Gamma_t' = \Gamma_t(\mathscr{s}'_{t\mathbf{C}_{t2}}; \mathscr{s}_{t\mathbf{J}_t\cdot}, \Sigma')$ is globally invertible as a function of $\mathscr{s}'_{t\mathbf{C}_{t2}} \in \mathbf{S}_{t2}(\mathscr{s}_{t\mathbf{J}_t\cdot})$ and therefore we can express $\mathscr{s}'_{t\mathbf{C}_{t2}}$ as a function of $\Gamma_t' \in \bar{\Theta}_\Gamma^t$, given $\mathscr{s}_{t\mathbf{J}_t\cdot}$ and Σ': $\mathscr{s}'_{t\mathbf{C}_{t2}} = \tilde{s}_{t\mathbf{C}_{t2}}(\Gamma_t'; \mathscr{s}_{t\mathbf{J}_t\cdot}, \Sigma')$. Then, given any $\Gamma' \in \cap_{t=1}^T \bar{\Theta}_\Gamma^t$ and $\Sigma' \in \Theta_\Sigma$, by plugging $\mathscr{s}'_{t\mathbf{C}_{t2}} = \tilde{s}_{t\mathbf{C}_{t2}}(\Gamma'; \mathscr{s}_{t\mathbf{J}_t\cdot}, \Sigma')$ into (61), we can express each δ_{tj}' from $\delta'_{t\mathbf{J}_t} = (\delta'_{tj})_{j\in\mathbf{J}_t}$ as a function of the observed marginals $\mathscr{s}_{t\mathbf{J}_t\cdot}$:

$$
\begin{aligned}
\delta_{tj}' &= s_{tj}^{-1}((\mathscr{s}_{tj\cdot} - \sum_{\mathbf{b}\in\mathbf{C}_{t2}: j\in\mathbf{b}} \mathscr{s}'_{t\mathbf{b}})_{j\in\mathbf{J}_t}, \mathscr{s}'_{t\mathbf{C}_{t2}}; \Sigma') \\
&= s_{tj}^{-1}((\mathscr{s}_{tj\cdot} - \sum_{\mathbf{b}\in\mathbf{C}_{t2}: j\in\mathbf{b}} \tilde{s}_{t\mathbf{b}}(\Gamma'; \mathscr{s}_{t\mathbf{J}_t\cdot}, \Sigma'))_{j\in\mathbf{J}_t}, \tilde{s}_{t\mathbf{C}_{t2}}(\Gamma'; \mathscr{s}_{t\mathbf{J}_t\cdot}, \Sigma'); \Sigma') \\
&= \mathscr{s}_{tj\cdot}^{-1}(\mathscr{s}_{t\mathbf{J}_t\cdot}; \Gamma', \Sigma')
\end{aligned}
$$

and determine the remaining $\delta'_{t\mathbf{b}}$ for each $\mathbf{b} \in \mathbf{C}_{t2}$ by $\delta'_{t\mathbf{b}} = \sum_{j\in\mathbf{b}} \delta'_{tj} + \Gamma'_{\mathbf{b}}$, so that $s_{t\mathbf{b}}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma') = \mathscr{s}'_{t\mathbf{b}}$ for each $\mathbf{b} \in \mathbf{C}_{t1}$. Then, for any $j \in \mathbf{J}_t$, we obtain $s_{tj\cdot}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma') = \mathscr{s}_{tj\cdot}$ and finally:

$$
s_{t\mathbf{J}_t\cdot}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma') = \mathscr{s}_{t\mathbf{J}_t\cdot}.
$$

This shows existence. To prove uniqueness, suppose that there exists another $\delta''_{t\mathbf{J}_t} \neq \delta'_{t\mathbf{J}_t}$ such that $s_{t\mathbf{J}_t\cdot}(\delta''_{t\mathbf{J}_t}; \Gamma', \Sigma') = \mathscr{s}_{t\mathbf{J}_t\cdot}$. Then, $\delta_t'' \neq \delta_t'$. Because Σ' and $\mathscr{s}_{t\mathbf{J}_t\cdot}$ are given, according to Lemma 1, there exists some $\mathbf{b} \in \mathbf{C}_{t2}$ for which $s_{t\mathbf{b}}(\delta''_{t\mathbf{J}_t}; \Gamma', \Sigma') \neq s_{t\mathbf{b}}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma')$. This contradicts the global invertibility of $\Gamma_t' \in \Gamma_t(\mathbf{S}_{t2}(\mathscr{s}_{t\mathbf{J}_t\cdot}); \mathscr{s}_{t\mathbf{J}_t\cdot}, \Sigma')$ with respect to $\mathscr{s}'_{t\mathbf{C}_{t2}} \in \mathbf{S}_{t2}(\mathscr{s}_{t\mathbf{J}_t\cdot})$ and completes the proof.

7.16 Proof of Theorem 8

We assume the same regularity conditions as in Theorem 6. Define

$$
\ell^c(\Gamma', \Sigma'; \mathscr{s}_1, ..., \mathscr{s}_T) = \ell((\delta_{t\mathbf{J}_t\cdot}(\mathscr{s}_{t\mathbf{J}_t\cdot}; \Gamma', \Sigma'))_{t=1,...,T}; \mathscr{s}_1, ..., \mathscr{s}_T).
$$

As for Theorem 6, to prove consistency we verify the following four conditions.

1. (Γ, Σ) **is the unique maximizer of** $\ell^c(\Gamma', \Sigma'; \mathscr{s}_1, ..., \mathscr{s}_T)$ **in** $\Theta_\Gamma \times \Theta_\Sigma$**.** Given Assumptions 1–3 and that the true $(\delta_{t\mathbf{J}_t})_{t\in\mathbf{T}\setminus\mathbf{T}_0} \in \Delta^{\mathrm{ID}}$, Lemma 3 guarantees that the true $(\delta_{1\mathbf{J}_1}, ..., \delta_{T\mathbf{T}}, \Gamma, \Sigma)$ is the unique maximizer of $\ell(\delta'_{1\mathbf{J}_1}, ..., \delta'_{T\mathbf{T}}, \Gamma', \Sigma')$ in Θ. Theorem 7 then implies that (Γ, Σ) is the unique maximizer of $\ell^c(\Gamma', \Sigma'; \mathscr{s}_1, ..., \mathscr{s}_T)$ in $\Theta_\Gamma \times \Theta_\Sigma$.

2. $\Theta_\Gamma \times \Theta_\Sigma$ **is compact.** This is guaranteed by the definition of Θ.

3. $\ell^c(\Gamma', \Sigma'; \mathscr{s}_1, ..., \mathscr{s}_T)$ **is continuous with respect to** (Γ', Σ') **in** $\Theta_\Gamma \times \Theta_\Sigma$**.** As shown in the proof of Theorem 6, $s_{t\mathbf{b}}(\delta'_t; F(\cdot; \Sigma'))$ is continuous with respect to δ'_t and Σ'. Consequently, the inverse market share function, $s_t^{-1}(\mathscr{s}'_t; \Sigma')$ is continuous with respect to $(\mathscr{s}'_t, \Sigma')$, and therefore continuous with respect to $((\mathscr{s}_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_{t2}}, \Sigma')$. Then, $\Gamma_t((\mathscr{s}_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_{t2}}, \Sigma')$, as defined in the proof of Theorem 7, is continuous with respect to $(\{\mathscr{s}_{t\mathbf{b}}\}_{\mathbf{b}\in\mathbf{C}_{t2}}, \Sigma')$. By applying the invertibility result from Theorem 7 and the continuous dependence with respect to Σ', we obtain that $\delta_{t\mathbf{J}_t\cdot}(\mathscr{s}_t; \Gamma', \Sigma')$ is continuous with respect to (Γ', Σ'). Combining this with the continuity of $\ell(\theta'; \mathscr{s}_1, ..., \mathscr{s}_T)$, we obtain the desired condition.

4. $\sup_{(\Gamma',\Sigma')\in\Theta_\Gamma\times\Theta_\Sigma} |\ell^c_I(\Gamma', \Sigma'; \hat{\mathscr{s}}_1, ..., \hat{\mathscr{s}}_T) - \ell^c(\Gamma', \Sigma')| \xrightarrow{p} 0$**.** This follows from:

$$\sup_{(\Gamma',\Sigma')\in\Theta_\Gamma\times\Theta_\Sigma} |\ell^c_I(\Gamma', \Sigma'; \hat{\mathscr{s}}_1, ..., \hat{\mathscr{s}}_T) - \ell^c(\Gamma', \Sigma')| \leq \sup_{\theta'\in\Theta} |\ell_I(\theta'; \hat{\mathscr{s}}_1, ..., \hat{\mathscr{s}}_T) - \ell(\theta')| \xrightarrow{p} 0. \tag{63}$$

According to Theorem 2.1 by Newey and McFadden (1994), the four conditions verified above guarantee the consistency of $(\hat{\Gamma}^c, \hat{\Sigma}^c)$. By applying the invertibility result from Theorem 7 and Slutsky's Theorem, $\hat{\delta}^c$ is also consistent. This completes the proof of consistency.

The proof of asymptotic normality is again based on Theorem 3.1 by Newey and McFadden (1994), according to which we need to verify the following six conditions.

1. $\hat{\theta}^c \xrightarrow{p} (\Gamma, \Sigma)$**.** This has just been shown above.

2. (Γ, Σ) **is an interior point of** $\Theta_\Gamma \times \Theta_\Sigma$**.** This is guaranteed by regularity condition 1.

3. $\ell^c_I(\Gamma', \Sigma'; \hat{\mathscr{s}}_1, ..., \hat{\mathscr{s}}_T)$ **is twice continuously differentiable in** $\Theta_\Gamma \times \Theta_\Sigma$**.** According to regularity condition 2, the market share function $s_{t\mathbf{b}}(\delta'_t; \Sigma')$, $t = 1, ..., T$ and $\mathbf{b} \in \mathbf{C}_t$, is twice continuously differentiable with respect to (δ'_t, Σ'), the inverse market share function $s_t^{-1}(\mathscr{s}'_t; \Sigma')$ is thus twice continuously differentiable with respect to $(\mathscr{s}'; \Sigma')$. By applying the invertibility result from Theorem 7, we obtain that $\delta_{t\mathbf{J}_t\cdot}(\mathscr{s}'_{t\mathbf{J}_t\cdot}; \Gamma', \Sigma')$ is twice continuously differentiable with respect to (Γ', Σ'). Because $\ell^c_I(\Gamma', \Sigma'; \hat{\mathscr{s}}_1, ..., \hat{\mathscr{s}}_T)$ is composed of $\ell_I(\delta_{t\mathbf{J}_t}, \Gamma', \Sigma')$ and $\delta_{t\mathbf{J}_t\cdot}(\mathscr{s}'_{t\mathbf{J}_t\cdot}; \Gamma', \Sigma')$, and both functions are twice continuously differentiable, $\ell^c_I(\Gamma', \Sigma'; \hat{\mathscr{s}}_1, ..., \hat{\mathscr{s}}_T)$ is also twice continuously differentiable with respect to (Γ', Σ').

4. $\sqrt{I}\frac{\partial \ell_I^c}{\partial(\Gamma',\Sigma')}\Big|_{(\Gamma',\Sigma')=(\Gamma,\Sigma)}$ **converges to a centered normal distribution.** We can write:

$$\begin{aligned}\frac{\partial \ell_I^c}{\partial(\Gamma',\Sigma')} &= \sum_{t=1,\dots,T} \frac{\partial \delta_{t\mathbf{J}_t}}{\partial(\Gamma',\Sigma')}\frac{\partial \ell_I}{\partial \delta_{t\mathbf{J}_t}} + \frac{\partial \ell_I}{\partial(\Gamma',\Sigma')} \\ &= \left[\left(\frac{\partial \delta_{t\mathbf{J}_t}}{\partial(\Gamma',\Sigma')}\right)_{t=1,\dots,T} \quad \mathbf{I}\right]\frac{\partial \ell_I}{\partial \theta'} \\ &= \left[\left(\frac{\partial \delta_{t\mathbf{J}_t}}{\partial(\Gamma',\Sigma')}\right)_{t=1,\dots,T} \quad \mathbf{I}\right]\sum_{t=1}^{T}\left(\left[\frac{\partial \log s_{t\mathbf{b}}}{\partial \theta'} - \frac{\partial \log s_{t0}}{\partial \theta'}\right]\Big|_{\theta'=\theta}\right)_{\mathbf{b}\in\mathbf{C}_{t1}} \sqrt{I}\left[\hat{\jmath}_t - \jmath_t\right].\end{aligned} \tag{64}$$

Evaluated at $(\Gamma',\Sigma')=(\Gamma,\Sigma)$, i.e., $\theta'=\theta$, $\sqrt{I}\frac{\partial \ell_I}{\partial \theta'}$ converges to a centered normal distribution. Then, $\frac{\partial \ell_I^c}{\partial(\Gamma',\Sigma')}\Big|_{(\Gamma',\Sigma')=(\Gamma,\Sigma)}$ also converges to a centered normal distribution.

5. $\sup_{(\Gamma',\Sigma')\in\Theta_\Gamma\times\Theta_\Sigma}\left|\frac{\partial^2 \ell_I^c}{\partial(\Gamma',\Sigma')^2}(\Gamma',\Sigma') - H^c(\Gamma',\Sigma')\right| \xrightarrow{p} 0$**.** Define

$$\begin{aligned}H^c(\Gamma',\Sigma') &= \frac{\partial^2 \ell(\{\delta_{t\mathbf{J}_t\cdot}(\Gamma',\Sigma')\}_{t=1,\dots,T};\Gamma',\Sigma',\jmath_1,\dots,\jmath_T)}{\partial(\Gamma',\Sigma')^2} \\ &= \sum_{t=1,\dots,T}\sum_{\mathbf{b}\in\mathbf{C}_t} \jmath_{t\mathbf{b}}\frac{\partial \log^2 s_{t\mathbf{b}}(\delta_{t\mathbf{J}_t\cdot}(\Gamma',\Sigma'),\Gamma',\Sigma')}{\partial(\Gamma',\Sigma')^2}.\end{aligned} \tag{65}$$

Then, under regularity condition 2, $H^c(\Gamma',\Sigma')$ is continuous in $\Theta_\Gamma\times\Theta_\Sigma$. Note that, similarly to (63), we have:

$$\begin{aligned}&\sup_{(\Gamma',\Sigma')\in\Theta_\Gamma\times\Theta_\Sigma}\left|\frac{\partial^2 \ell_I^c}{\partial(\Gamma',\Sigma')^2}(\Gamma',\Sigma') - H^c(\Gamma',\Sigma')\right| \\ \leq &\sum_{t=1,\dots,T,\mathbf{b}\in\mathbf{C}_t}\sup_{(\Gamma',\Sigma')\in\Theta_\Gamma\times\Theta_\Sigma}\left|\frac{\partial \log^2 s_{t\mathbf{b}}(\delta_{t\mathbf{J}_t\cdot}(\Gamma',\Sigma'),\Gamma',\Sigma')}{\partial(\Gamma',\Sigma')^2}\right| |\hat{\jmath}_{t\mathbf{b}} - \jmath_{t\mathbf{b}}|.\end{aligned} \tag{66}$$

Due to the twice continuous differentiability of $s_{t\mathbf{b}}(\theta_t';\Sigma')$ and $\delta_{t\mathbf{J}_t\cdot}(\jmath_{t\mathbf{J}_t\cdot};\Gamma',\Sigma')$ in the compact set $\Theta_\Gamma\times\Theta_\Sigma$, for $t=1,\dots,T$ and $\mathbf{b}\in\mathbf{C}_t$, $\sup_{(\Gamma',\Sigma')\in\Theta_\Gamma\times\Theta_\Sigma}\left|\frac{\partial \log^2 s_{t\mathbf{b}}(\delta_{t\mathbf{J}_t\cdot}(\Gamma',\Sigma'),\Gamma',\Sigma')}{\partial(\Gamma',\Sigma')^2}\right| < \infty$. Because $\hat{\jmath}_t \xrightarrow{p} \jmath_t$ for $t=1,\dots,T$, the right-hand side of (66) converges to zero in probability and $\sup_{(\Gamma',\Sigma')\in\Theta_\Gamma\times\Theta_\Sigma}\left|\frac{\partial^2 \ell_I^c}{\partial(\Gamma',\Sigma')^2}(\Gamma',\Sigma') - H^c(\Gamma',\Sigma')\right| \xrightarrow{p} 0$.

6. $H^c(\Gamma,\Sigma)=\frac{\partial^2\ell^c(\Gamma',\Sigma')}{\partial(\Gamma',\Sigma')^2}\Big|_{(\Gamma',\Sigma')=(\Gamma,\Sigma)}$ **is non-singular.** Note that

$$\begin{aligned}
H^c(\Gamma',\Sigma') &= \frac{\partial^2\ell(\{\delta_{t\mathbf{J}_t\cdot}(\Gamma',\Sigma')\}_{t=1,\dots,T};\Gamma',\Sigma';\jmath_1,\dots,\jmath_T)}{\partial(\Gamma',\Sigma')^2}\\
&=\sum_{t=1,..,T}\left[\frac{\partial\delta_{t\mathbf{J}_t\cdot}}{\partial(\Gamma',\Sigma')}\frac{\partial^2\ell}{\partial\delta'^2_{t\mathbf{J}_t}}\left(\frac{\partial\delta_{t\mathbf{J}_t\cdot}}{\partial(\Gamma',\Sigma')}\right)^{\mathrm{T}}+\frac{\partial\delta_{t\mathbf{J}_t\cdot}}{\partial(\Gamma',\Sigma')}\frac{\partial^2\ell}{\partial\delta'_{t\mathbf{J}_t}\partial(\Gamma',\Sigma')}\right]\\
&+\sum_{t=1,..,T}\sum_{j\in\mathbf{J}_t}\frac{\partial\ell}{\partial\delta'_{tj}}\frac{\partial^2\delta_{tj\cdot}}{\partial(\Gamma',\Sigma')^2}\\
&+\sum_{t=1,..,T}\frac{\partial\delta_{t\mathbf{J}_t\cdot}}{\partial(\Gamma',\Sigma')}\frac{\partial^2\ell}{\partial(\Gamma',\Sigma')\partial\delta'_{t\mathbf{J}_t}}+\frac{\partial^2\ell}{\partial(\Gamma',\Sigma')^2}.
\end{aligned}$$

At $(\Gamma',\Sigma')=(\Gamma,\Sigma)$, $\delta'_{t\mathbf{J}_t}=\delta_{t\mathbf{J}_t\cdot}(\Gamma,\Sigma)$ and $\frac{\partial\ell}{\partial\delta'_{t\mathbf{J}_t}}=0$. Then,

$$H^c(\Gamma,\Sigma)=\left[\left(\frac{\partial\delta_{t\mathbf{J}_t}}{\partial(\Gamma',\Sigma')}\right)_{t=1,\dots,T}\quad\mathbf{I}\right]H(\delta)\left[\left(\frac{\partial\delta_{t\mathbf{J}_t}}{\partial(\Gamma',\Sigma')}\right)_{t=1,\dots,T}\quad\mathbf{I}\right]^{\mathrm{T}}.$$

Because $\left[\left\{\frac{\partial\delta_{t\mathbf{J}_t}}{\partial(\Gamma',\Sigma')}\right\}_{t=1,\dots,T}\quad\mathbf{I}\right]$ is of full row rank and $H(\delta)$ is non-singular, $H^c(\Gamma,\Sigma)$ is therefore non-singular.

All the six conditions of Theorem 3.1 by Newey and McFadden (1994) are satisfied and there exists W_2 such that $\sqrt{I}[(\hat\Gamma^c,\hat\Sigma^c)-(\Gamma,\Sigma)]\xrightarrow{d}\mathcal{N}(0,W_2)$. By applying the invertibility result from Theorem 7, we have:

$$\begin{aligned}
\sqrt{I}(\hat\delta^c_{t\mathbf{J}_t}-\delta_{t\mathbf{J}_t}) &= \sqrt{I}(\tilde\delta_{t\mathbf{J}_t}(\hat\jmath,\hat\Gamma^c,\hat\Sigma^c)-\tilde\delta_{t\mathbf{J}_t}(\hat\jmath,\Gamma,\Sigma)+\tilde\delta_{t\mathbf{J}_t}(\hat\jmath,\Gamma,\Sigma)-\tilde\delta_{t\mathbf{J}_t}(\jmath,\Gamma,\Sigma))\\
&=\frac{\partial\tilde\delta_{t\mathbf{J}_t}(\hat\jmath;\Gamma',\Sigma')}{\partial(\Gamma',\Sigma')}\Big|_{(\Gamma',\Sigma')=(\tilde\Gamma,\tilde\Sigma)}\sqrt{I}[(\hat\Gamma^c,\hat\Sigma^c)-(\Gamma,\Sigma)]\\
&+\frac{\partial\tilde\delta_{t\mathbf{J}_t}(\jmath';\Gamma,\Sigma)}{\partial\jmath'}\Big|_{\jmath'=\tilde\jmath}\sqrt{I}(\hat\jmath-\jmath).
\end{aligned}$$

Using the following Taylor expansion of $\frac{\partial\ell^c_I}{\partial(\Gamma',\Sigma')}$ around (Γ,Σ):

$$0=\frac{\partial\ell^c_I}{\partial(\Gamma',\Sigma')}\Big|_{(\Gamma',\Sigma')=(\hat\Gamma^c,\hat\Sigma^c)}=\frac{\partial\ell^c_I}{\partial(\Gamma',\Sigma')}\Big|_{(\Gamma',\Sigma')=(\Gamma,\Sigma)}+\frac{\partial^2\ell^c_I}{\partial(\Gamma',\Sigma')^2}\Big|_{(\Gamma',\Sigma')=(\tilde\Gamma,\tilde\Sigma)}[(\hat\Gamma^c,\hat\Sigma^c)-(\Gamma,\Sigma)],$$

we obtain

$$\begin{aligned}
\sqrt{I}[(\hat{\Gamma}^c,\hat{\Sigma}^c)-(\Gamma,\Sigma)] &= -\left[\frac{\partial^2 \ell_I^c}{\partial(\Gamma',\Sigma')^2}\Big|_{(\Gamma',\Sigma')=(\tilde{\Gamma},\tilde{\Sigma})}\right]^{-1}\sqrt{I}\frac{\partial \ell_I^c}{\partial(\Gamma',\Sigma')}\Big|_{(\Gamma',\Sigma')=(\Gamma,\Sigma)} \\
&= -\left[\frac{\partial^2 \ell_I^c}{\partial(\Gamma',\Sigma')^2}\Big|_{(\Gamma',\Sigma')=(\tilde{\Gamma},\tilde{\Sigma})}\right]^{-1} \\
&\quad \left[\left(\frac{\partial \delta_{t\mathbf{J}_t}}{\partial(\Gamma',\Sigma')}\right)_{t=1,\ldots,T} \quad \mathbf{I}\right]\sum_{t=1}^{T}\left(\left[\frac{\partial \log s_{t\mathbf{b}}}{\partial\theta'}-\frac{\partial \log s_{t0}}{\partial\theta'}\right]\Big|_{\theta'=\theta}\right)_{\mathbf{b}\in\mathbf{C}_{t1}}\sqrt{I}\left[\hat{\jmath}_t-\jmath_t\right].
\end{aligned}$$

Since $\sqrt{I}(\hat{\jmath}-\jmath)$ converges to a centered normal distribution, by using Slutsky's Theorem and the consistency of $\hat{\delta}^c_{t\mathbf{J}_t}$ and $(\hat{\Gamma}^c,\hat{\Sigma}^c)$, we conclude that $\sqrt{I}(\hat{\delta}^c_{t\mathbf{J}_t}-\delta_{t\mathbf{J}_t})$ converges to a centered normal distribution. This completes the proof.

7.17 Proof of Proposition 1

Because producer f only owns two brands, j and j', the profit from choosing prices $p_f = (p_j, p_{j'}, p_{(j,j')})$ simplifies from (19) to:

$$\begin{aligned}
\pi_f(p_f,p_{-f}) &= \pi_f^1(p_f,p_{-f})+\pi_f^2(p_f,p_{-f})+\pi_f^3(p_f,p_{-f}), \text{ with} \\
\pi_f^1 &= s_j(p_f,p_{-f})(p_j-c_j)+s_{j'}(p_f,p_{-f})(p_{j'}-c_{j'}), \\
\pi_f^2 &= s_{(j,j')}(p_f,p_{-f})(p_{(j,j')}-c_j-c_{j'}), \\
\pi_f^3 &= s_{j,-f}(p_f,p_{-f})(p_j-c_j)+s_{j',-f}(p_f,p_{-f})(p_{j'}-c_{j'}).
\end{aligned} \tag{67}$$

Denote $\pi_f^1+\pi_f^3$ by $\pi_f^{1,3}$. Then, the FOCs for the maximization of (67) with respect to p_j under pure components pricing are:

$$\frac{\partial \pi_f^{1,3}}{\partial p_j}+\frac{\partial \pi_f^{1,3}}{\partial p_{(j,j')}}+\frac{\partial \pi_f^{2}}{\partial p_j}+\frac{\partial \pi_f^{2}}{\partial p_{(j,j')}}=0. \tag{68}$$

Denote by $p_f^* = (p_j^*, p_{j'}^*, p_j^*+p_{j'}^*)$ a vector of pure components prices that satisfies (68) and the analogous equation for $p_{j'}$. Differently, for the case of mixed bundling, the FOCs for p_j and for $p_{(j,j')}$ are:

$$\begin{aligned}
&\text{For } j: \quad \frac{\partial \pi_f^{1,3}}{\partial p_j}+\frac{\partial \pi_f^{2}}{\partial p_j}=0. \\
&\text{For } (j,j'): \quad \frac{\partial \pi_f^{1,3}}{\partial p_{(j,j')}}+\frac{\partial \pi_f^{2}}{\partial p_{(j,j')}}=0.
\end{aligned} \tag{69}$$

In what follows, we establish the signs of the left-hand sides of (69) when evaluated at p_f^*. By Slutsky symmetry, $\frac{\partial(s_j+s_{j,-f})}{\partial p_{(j,j')}} = \frac{\partial s_{(j,j')}}{\partial p_j}$ and $\frac{\partial(s_{j'}+s_{j',-f})}{\partial p_{(j,j')}} = \frac{\partial s_{(j,j')}}{\partial p_{j'}}$. Then,

$$\begin{aligned}\frac{\partial \pi_f^{1,3}}{\partial p_{(j,j')}} &= \frac{\partial(s_j + s_{j,-f})}{\partial p_{(j,j')}}(p_j^* - c_j) + \frac{\partial(s_{j'} + s_{j',-f})}{\partial p_{(j,j')}}(p_{j'}^* - c_{j'}) \\ &= \frac{\partial s_{(j,j')}}{\partial p_j}(p_j^* - c_j) + \frac{\partial s_{(j,j')}}{\partial p_{j'}}(p_{j'}^* - c_{j'}).\end{aligned} \tag{70}$$

Hence, the left-hand side of (69) for (j, j') evaluated at p_f^* can be expressed as:

$$\begin{aligned}\frac{\partial \pi_f^{1,3}}{\partial p_{(j,j')}} + \frac{\partial \pi_f^{2}}{\partial p_{(j,j')}} &= \frac{\partial s_{(j,j')}}{\partial p_j}(p_j^* - c_j) + \frac{\partial s_{(j,j')}}{\partial p_{j'}}(p_{j'}^* - c_{j'}) + \frac{\partial s_{(j,j')}}{\partial p_{(j,j')}}(p_j^* + p_{j'}^* - c_j - c_{j'}) + s_{(j,j')} \\ &= s_{(j,j')}\left[\mathscr{E}_j^{(j,j')}\lambda_j + \mathscr{E}_{j'}^{(j,j')}\lambda_{j'} + \mathscr{E}_{(j,j')}^{(j,j')}\lambda_{(j,j')} + 1\right],\end{aligned} \tag{71}$$

where $\mathscr{E}_r^{(j,j')}$ is the elasticity of $s_{(j,j')}$ with respect to p_r evaluated at p_f^* and $\lambda_r = (p_r^* - c_r)/p_r^*$ is r's Lerner index, for $r \in \{j, j', (j, j')\}$. Note that $\mathscr{E}_j^{(j,j')}$ and $\mathscr{E}_{j'}^{(j,j')}$ are positive, while $\mathscr{E}_{(j,j')}^{(j,j')}$ is negative. Moreover, because of the mixed logit specification, $\mathscr{E}_j^{(j,j')}\lambda_j + \mathscr{E}_{j'}^{(j,j')}\lambda_{j'} + \mathscr{E}_{(j,j')}^{(j,j')}\lambda_{(j,j')}$ is always negative.[46] Consequently, the sign of (71) is determined by:

$$\mathscr{E}_j^{(j,j')}\lambda_j + \mathscr{E}_{j'}^{(j,j')}\lambda_{j'} + \mathscr{E}_{(j,j')}^{(j,j')}\lambda_{(j,j')} < -1. \tag{73}$$

When condition (73) is satisfied, the derivative of profit with respect to $p_{(j,j')}$ evaluated at p_f^* is negative and f will have an incentive to lower $p_{(j,j')}$ relative to $p_{(j,j')}^* = p_j^* + p_{j'}^*$. Symmetrically, because of (68), the derivatives of profit with respect to p_j and $p_{j'}$ evaluated at p_f^* are positive and f will have an incentive to increase both p_j and $p_{j'}$ relative to p_j^* and $p_{j'}^*$.

[46]It is sufficient to show that $\frac{\partial s_{(j,j')}}{\partial p_j}(p_j^* - c_j) + \frac{\partial s_{(j,j')}}{\partial p_{j'}}(p_{j'}^* - c_{j'}) + \frac{\partial s_{(j,j')}}{\partial p_{(j,j')}}(p_j^* + p_{j'}^* - c_j - c_{j'}) < 0$. For simplicity of exposition, assume a mixed logit model with household i-specific price coefficient α_i^p. Suppose that $\alpha_i^p < 0$ for all households, and that the price-cost margins of both brands ($p_j^* - c_j$ and $p_{j'}^* - c_{j'}$) are positive, then:

$$\begin{aligned}&\frac{\partial s_{(j,j')}}{\partial p_j}(p_j^* - c_j) + \frac{\partial s_{(j,j')}}{\partial p_{j'}}(p_{j'}^* - c_{j'}) + \frac{\partial s_{(j,j')}}{\partial p_{(j,j')}}(p_j^* + p_{j'}^* - c_j - c_{j'}) \\ &= \left[\frac{\partial s_{(j,j')}}{\partial p_j} + \frac{\partial s_{(j,j')}}{\partial p_{(j,j')}}\right](p_j^* - c_j) + \left[\frac{\partial s_{(j,j')}}{\partial p_{j'}} + \frac{\partial s_{(j,j')}}{\partial p_{(j,j')}}\right](p_{j'}^* - c_{j'}) \\ &= \int \alpha_i^p s_{i(j,j')}\left[-(s_{ij} + s_{i(j,-f)}) + (1 - s_{i(j,j')})\right] dF(\mu_i)(p_j^* - c_j) \\ &+ \int \alpha_i^p s_{i(j,j')}\left[-(s_{ij'} + s_{i(j',-f)}) + (1 - s_{i(j,j')})\right] dF(\mu_i)(p_{j'}^* - c_{j'}) \\ &= \int \alpha_i^p s_{i(j,j')}\left[1 - s_{i(j,\cdot)}\right] dF(\mu_i)(p_j^* - c_j) + \int \alpha_i^p s_{i(j,j')}\left[1 - s_{i(j',\cdot)}\right] dF(\mu_i)(p_{j'}^* - c_{j'}) \\ &< 0.\end{aligned} \tag{72}$$