DISCUSSION PAPER SERIES

DP14363

IDENTIFICATION AND ESTIMATION OF DEMAND FOR BUNDLES

Alessandro Iaria and Ao Wang

INDUSTRIAL ORGANIZATION

ISSN 0265-8003

IDENTIFICATION AND ESTIMATION OF DEMAND FOR BUNDLES

Alessandro Iaria and Ao Wang

Discussion Paper DP14363
Published 29 January 2020
Submitted 28 January 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

IDENTIFICATION AND ESTIMATION OF DEMAND FOR BUNDLES

Abstract

We present novel identification and estimation results for a mixed logit model of demand for bundles with endogenous prices given bundle-level market shares. Our approach hinges on an affine relationship between the utilities of single products and of bundles, on an essential real analytic property of the mixed logit model, and on the existence of exogenous cost shifters. We propose a new demand inverse in the presence of complementarity that enables to concentrate out of the likelihood function the (potentially numerous) market-product specific average utilities, substantially alleviating the challenge of dimensionality inherent in estimation. To illustrate the use of our methods, we estimate demand and supply in the US ready-to-eat cereal industry, where the proposed MLE reduces the numerical search from approximately 12000 to 130 parameters. Our estimates suggest that ignoring Hicksian complementarity among different products often purchased in bundles may result in misleading demand estimates and counterfactuals.

JEL Classification: N/A

Keywords: N/A

Alessandro Iaria - iaria.alessandro@gmail.com
University of Bristol and CEPR

Ao Wang - ao.wang@ensae.fr
CREST

Acknowledgements
Iaria gratefully acknowledges funding from the British Academy/Leverhulme SRG no. 170142. Very special thanks to Xavier D'Haultfoeuille for the many discussions and suggestions at each stage of this and related projects. We are extremely grateful for the insightful suggestions to Abi Adams, Greg Crawford, Michaela Draganska, Gautam Gowrisankaran, Kyoo il Kim, Chris Murris, David Pacini, Ariel Pakes, Andrea Pozzi, Pasquale Schiraldi, Philipp Schmidt-Dengler, Howard Smith, Sami Stouli, Frank Verboven, Ali Yurukoglu, and to all seminar participants at the Bristol Researh Day, 11th Berlin IO Day, 20th CEPR Applied IO (Madrid), CIREQ Montreal, 8th EIEF-Unibo-IGIER Bocconi IO Workshop, 2019 European Winter Meeting of the Econometric Society (Rotterdam), 2019 GSE Summer Forum Applied IO, Guanghua School of Management, 17th IIOC (Boston), IWIIE (Nanjing), 2019 JEI (Madrid), 29th Jerusalem Summer School of Economics (IO), Oxford, and PSE. 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.

Identification and Estimation of Demand for Bundles

Alessandro Iaria and Ao Wang*

This Version: January 2020
First Version: March 2019

Abstract

We present novel identification and estimation results for a mixed logit model of demand for bundles with endogenous prices given bundle-level market shares. Our approach hinges on an affine relationship between the utilities of single products and of bundles, on an essential real analytic property of the mixed logit model, and on the existence of exogenous cost shifters. We propose a new demand inverse in the presence of complementarity that enables to concentrate out of the likelihood function the (potentially numerous) market-product specific average utilities, substantially alleviating the challenge of dimensionality inherent in estimation. To illustrate the use of our methods, we estimate demand and supply in the US ready-to-eat cereal industry, where the proposed MLE reduces the numerical search from approximately 12000 to 130 parameters. Our estimates suggest that ignoring Hicksian complementarity among different products often purchased in bundles may result in misleading demand estimates and counterfactuals.

*University of Bristol and CEPR (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'Haultfœuille for the many discussions and suggestions at each stage of this and related projects. We are extremely grateful for the insightful suggestions to Abi Adams, Greg Crawford, Michaela Draganska, Gautam Gowrisankaran, Kyoo il Kim, Chris Murris, David Pacini, Ariel Pakes, Andrea Pozzi, Pasquale Schiraldi, Philipp Schmidt-Dengler, Howard Smith, Sami Stouli, Frank Verboven, Ali Yurukoglu, and to all seminar participants at the Bristol Researh Day, 11th Berlin IO Day, 20th CEPR Applied IO (Madrid), CIREQ Montreal, 8th EIEF-Unibo-IGIER Bocconi IO Workshop, 2019 European Winter Meeting—Econometric Society (Rotterdam), 2019 GSE Summer Forum Applied IO, Guanghua School of Management, 17th IIOC (Boston), IWIIE (Nanjing), 2019 JEI (Madrid), 29th Jerusalem Summer School of Economics (IO), Oxford, and PSE. 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 single products (Debreu (1959), Varian (1992), and Mas-Colell et al. (1995)), allowing for both substitutability and complementarity. Despite important exceptions (Manski and Sherman (1980), Hendel (1999), Dubé (2004), Gentzkow (2007), and Thomassen et al. (2017)), the models routinely used to estimate demand rely on the assumption that each of the products purchased in a bundle is chosen independently, precluding the possibility of complementarity and potentially leading to incorrect estimates and counterfactuals.

Models of demand for bundles face non-trivial identification challenges (Gentzkow, 2007), even in settings with a limited number of products (Fox and Lazzati (2017) and Allen and Rehbeck (2019)). 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 (Berry et al., 2014). These difficulties forced empirical researchers either to focus on applications with a limited number of products (typically two or three) or to make restrictive assumptions on the parameters capturing potential synergies among the products within bundles (typically a common parameter for all bundles and individuals).[1]

We tackle these challenges and propose empirical methods that are practically useful in applications with more than a few products. In particular, we study the identification and estimation of a mixed logit model of demand for bundles with endogenous prices given observations on bundle-level market shares. Our arguments hinge on the affine relationship between the utilities of single products and of bundles typical of models along the lines of Gentzkow (2007)'s: the average utility of any bundle equals the sum of the average utilities of the single products plus an extra term capturing their potential demand synergies. This utility structure allows *(i)* for a novel identification approach based on the existence of exogenous but potentially unobserved cost shifters and *(ii)* to alleviate the challenge of dimensionality in estimation by means of a new demand inverse in the presence of complementarity.

Our approach is based on a symmetry assumption about the *average* demand synergies across markets: while the demand synergies for any specific bundle may be unobserved and heterogeneous across individuals, their average is required to be constant across markets with the same observable characteristics (e.g., demographics and prices). Importantly, we propose a specification test for this symmetry assumption on the basis of partial identification methods that can be performed prior to the estimation of the full model. Under this symmetry assumption and regularity conditions similar to Rothenberg (1971), we derive necessary and sufficient rank conditions for the local identification of the model with endogenous prices. This result formalizes Gentzkow (2007)'s insight that, when the average demand synergies are "similar"

[1]Throughout the paper, we refer to the parameters capturing the potential synergies among products within bundles simply as *demand synergies* or *demand synergy parameters*.

across markets, the availability of data on many markets will help identification. The "necessity" part of the result is informative about the limits of identification in models of demand for bundles: the separate identification of demand synergies and of the distribution of random coefficients is not immediate, and one needs observations on "enough" markets to achieve it.

We provide novel sufficient conditions for the global identification of the model with endogenous prices to hold almost everywhere. Our argument combines three main ingredients: a finite number of elements in the identification set, an essential real analytic property of the mixed logit model, and the existence of exogenous cost shifters. We assume that the identification set does not have infinitely many elements and, building on Chernozhukov et al. (2007) and on Romano and Shaikh (2012), propose testable conditions to verify this in practice. We show that the mixed logit market share function is real analytic with respect to the market-product specific average utilities. This further shrinks the identification set in the presence of exogenous variation in the market-product specific average utilities. We then demonstrate that cost shifters can provide the required exogenous variation when the endogenous prices are generated by a large class of pure components and mixed bundling price-setting models.[2] We finally attain global identification almost everywhere by assuming the existence of exogenous cost shifters that are potentially unobserved but identifiable from observed market shares and prices. One can then interpret our identification strategy as based on the existence of "unobserved" but "identifiable" instruments, the exogenous cost shifters.

We propose a 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 et al. (2015), Grzybowski and Verboven (2016), Ruiz et al. (2017), and Ershov et al. (2018)). 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 can be too large to be handled numerically (Berry et al., 2014)). We tackle this practical bottleneck by a novel demand inverse designed to handle complementarity among products in models along the lines of Gentzkow (2007)'s. For any given value of the other parameters, we establish a one-to-one mapping between the observed product-level market shares and the market-product specific average utilities.[3] This enables to concentrate out of the likelihood the potentially large number of market-product specific average utilities and to substantially

[2] For classic treatments of pure components and mixed bundling pricing strategies, see Adams and Yellen (1976), Lewbel (1985), McAfee et al. (1989). For more recent contributions, see Armstrong and Vickers (2010), Chu et al. (2011), Armstrong (2013), and Zhou (2017).

[3] Demand inverses at the *bundle*-level can simply rely on the classic results by Berry (1994) and Berry et al. (2013) as long as the *bundles* in the demand system are substitutes. However, if some of the *products* are complements, these classic results do not imply the invertibility of the demand system at the *product*-level. Our product-level demand inverse is instead based on the P-matrix property of Gale and Nikaido (1965), which does not require the products to be substitutes.

simplify the MLE's numerical search: in our application, the numerical search is reduced from approximately 12000 to 130 parameters. We show that our assumptions for global identification guarantee consistency and asymptotic normality of this estimator.

We illustrate our methods in the context of the ready-to-eat (RTE) cereal industry in the USA. We revisit the classic studies by Nevo (2000, 2001), and allow for Hicksian complementarity among different RTE cereal brands in demand estimation.[4] The households in our data are observed to purchase two or more different brands of RTE cereals in approximately 20% of their shopping trips. Our data record purchases rather than consumption: the purchases of different RTE cereal brands during the same shopping trip can clearly be motivated beyond synergies in consumption. For example, if households face shopping costs for each visit to a store, one-stop shopping may be preferred to multi-stop shopping (Pozzi (2012) and Thomassen et al. (2017)). Moreover, if households delegate grocery shopping to one person, then preference for variety may lead to the purchase of multiple brands on any shopping trip to accommodate the different needs of the household (Hendel (1999) and Dubé (2004)).

Our model encompasses these alternative mechanisms: the demand synergies are catch-all parameters that may reflect, for example, synergies in consumption, shopping costs, and preference for variety. We try to distinguish empirically the contribution of some of these possible mechanisms to the estimated demand synergies. Our results show that demand for RTE cereals exhibits substantial Hicksian complementarity and that around 75% of it does not seem to be explained by shopping costs or by preference for variety. We compare our estimation results from the full model to those from a model of demand for single brands (similar to Nevo (2000, 2001)) and show that ignoring Hicksian complementarity may result in misleading demand estimates and counterfactuals (see also Fosgerau et al. (2019)). In particular, estimates from the full model support the classic Cournot (1838)'s insight that, in the presence of Hicksian complementarity, mergers can be welfare enhancing; while those from a standard model that does not allow for it predict that mergers are detrimental for consumer surplus.

Related Literature. There is a growing empirical literature leveraging the estimation of demand for bundles. Manski and Sherman (1980) study households' choices of motor vehicle holdings; Hendel (1999) studies preference for variety for personal computers, while Dubé (2004) and Chan (2006) for soft carbonated drinks; Nevo et al. (2005) study the decision of libraries to subscribe to economics and business journals; Gentzkow (2007) and Gentzkow et al. (2014) investigate competition and complementarity among newspapers; Augereau et al. (2006) the returns from adoption of technological standards; Liu et al. (2010) and Grzybowski and Verboven (2016) the complementarity among telecommunication services; Crawford and Yu-

[4] Following Samuelson (1974) and Gentzkow (2007), we rely on the classic Hicksian notion of complementarity: we consider two brands as complements whenever their cross-price elasticity of (compensated) demand is negative. For recent discussions on complementarity in empirical models of demand, see Manzini et al. (2018), Dubé (2019), and Iaria and Wang (2019).

rukoglu (2012) and Crawford et al. (2018) the problem of bundling and vertical restraints in cable television, while Ho et al. (2012) in the video rental industry; Kretschmer et al. (2012) study the adoption of complementary innovations; Lee et al. (2013) the complementarity between milk and RTE cereals; Song et al. (2017) the relationship between mergers and inter-firm bundling in the pharmaceutical industry; Ruiz et al. (2017) propose a machine learning model of demand for bundles, Thomassen et al. (2017) study the problem of transportation costs in grocery shopping; Ershov et al. (2018) the complementarity between potato chips and soft carbonated drinks; and Fosgerau et al. (2019) the complementarity between different brands of RTE cereals. We add to this empirical literature by providing novel identification and estimation methods for models along the lines of Gentzkow (2007)'s, specifically accounting for price endogeneity and alleviating the challenge of dimensionality inherent in estimation.

The global identification of non-linear models is notoriously complex to demonstrate (Newey and McFadden (1994) and Lewbel (2019)). Researchers typically resume to non-verifiable abstract conditions (Rothenberg (1971), Bowden (1973), and Komunjer (2012)) or focus on weaker identification concepts altogether, such as local identification (Rothenberg (1971), Sargan (1983), and Lewbel (2012)) or partial identification (Manski (1989), Manski (2003), and Chesher and Rosen (2017)). We contribute to this literature by providing sufficient conditions for global identification that are testable (on the basis of partial identification methods), rooted in economic theory (to address price endogeneity), and weaker than the classics (Rothenberg (1971), Bowden (1973), and Komunjer (2012)). The relative advantage of our conditions follows from a real analytic property we show to be satisfied by mixed logit models given any distribution of random coefficients (parametric or non-parametric), which allows us to relax the strict concavity of the likelihood function (or similar criterion functions). Fox et al. (2012) and il Kim (2014) also exploit the real analytic properties of logit models to achieve global identification, but in more restrictive frameworks. il Kim (2014) shows the real analytic property for multinomial logit and for nested logit models, while Fox et al. (2012) show it for mixed logit models with random coefficients defined over compact supports—thereby ruling out, for example, normal and the log-normal distributions.

In the context of identification of models of demand for bundles, we add to the discussions by, for example, Fox and Lazzati (2017) and Allen and Rehbeck (2019). Fox and Lazzati (2017) propose sufficient conditions for the non-parametric identification of demand for bundles (and binary games of complete information) on the basis of additively separable excluded regressors. Allen and Rehbeck (2019) instead study the non-parametric identification of a large class of demand models, among which demand for bundles, by exploiting variation in the substitution and complementarity patterns among different products. While these papers make fewer distributional assumptions and can be preferred in situations with small choice sets and exogenous regressors, our arguments apply more readily to cases with larger choice sets, endogenous prices, and in general lead to practically convenient estimators.

Our mixed logit model of demand for bundles can be seen as a special case of the general non-parametric framework by Berry and Haile (2014). Berry and Haile (2014)' identification argument relies on the availability of *observed* instruments both to pin down the distribution of random coefficients and to address price endogeneity. In contrast, Gentzkow (2007)'s utility structure allows us to propose a complementary identification strategy based on *unobserved* instruments: we rely on the existence of "unobserved" but "identifiable" cost shifters and on conditional symmetry restrictions among the average demand synergies across markets. While less general in abstract terms, our arguments are more applicable to cases with limited observability of instruments and give rise to sizeable computational advantages in estimation.[5]

Our estimator contributes to the modern literature on the estimation of demand systems started by Berry et al. (1995) (henceforth BLP). For example, Berry et al. (2004b), Freyberger (2015), and Armstrong (2016b) investigate the asymptotic properties of GMM estimators of demand systems with endogenous prices. While these GMM estimators are more widely applicable provided the availability of observable instruments, in the context of demand for bundles our MLE represents a numerically convenient alternative in which the instruments need to exist but do not need to be observed. More recently, Compiani (2019) proposes a non-parametric estimator of demand models that accommodates complementarity among products. There is a trade-off between our proposed estimator and Compiani (2019)'s. His non-parametric estimator is more flexible than ours, but it is subject to a curse of dimensionality that may constrain its applicability to settings with small choice sets. Our MLE is less affected by dimensionality and can be implemented with larger choice sets.

Since Berry (1994), the identification and the estimation of demand systems with endogenous prices has been relying on the ability to "invert" market share equations to uniquely determine the implied product-specific average utilities—the so called demand inverse. A standard requirement for the invertibility of demand systems is for the products to be substitutes, see Berry et al. (2013). This requirement can be problematic in contexts with complementary products: for example, in a model of demand for bundles of newspapers, Fan (2013) rules out by assumption any complementarity in order to rely on the classic demand inverse by Berry (1994) at the newspaper-level. Our novel demand inverse addresses this issue and allows to invert product-level market share equations in the presence of complementarity.

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 our demand inverse and a related MLE. In section 6, we explore the practical relevance of our methods with an empirical illustration. In section 7, we conclude the paper with some final

[5]The classic identification argument based on observed instruments requires the performance of high-dimensional demand inverses at the bundle-level, while our argument based on *un*observed instruments allows for the performance of demand inverses only at the product-level. In practice, this implies the numerical inversion of a lower-dimensional demand system and leads to large computational advantages.

remarks. In appendix section 8, we report all the proofs and additional results.

2 Model and Notation

Imagine a cross-section of T independent markets denoted by $\mathbf{T}$, where each market $t \in \mathbf{T}$ is populated by $i = 1, ..., I$ individuals. Individual i in market t makes purchases exclusively in market t and is a different person from individual i in any other market $t' \neq t$. 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 $\mathbf{C}_t = \mathbf{C}_{t1} \cup \{0\}$ be the choice set specific to market t, which includes: the collection of "inside" options $\mathbf{C}_{t1}$ and the "outside" option $j = 0$ (i.e., the option not to purchase any product). In turn, the collection of inside options is defined as $\mathbf{C}_{t1} = \mathbf{J}_t \cup \mathbf{C}_{t2}$, where $\mathbf{C}_{t2}$ denotes the set of market-specific bundles of products. The set of all available bundles across all markets is $\mathbf{C}_2 = \cup_{t=1}^{T} \mathbf{C}_{t2}$. We refer to the cardinality of these sets as: $C_t = |\mathbf{C}_t|$, $C_{t1} = |\mathbf{C}_{t1}|$, $C_{t2} = |\mathbf{C}_{t2}|$, and $C_2 = |\mathbf{C}_2|$. We denote by $\mathbf{b}$ any element of the choice set $\mathbf{C}_t$, whereby some abuse of notation $\mathbf{b}$ may refer to a bundle, a single product, or the outside option.

The indirect utility of individual i in market t from purchasing 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 the market t-specific average utility of product $j \in \mathbf{J}_t$, μ_{itj} is an unobserved individual-specific utility deviation from δ_{tj}, while ε_{itj} and ε_{it0} are error terms. Throughout the paper, we treat the market t-specific average utilities as parameters to be identified and estimated. One can however reduce the number of parameters by using observable characteristics and making additional functional form assumptions.[6]

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

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

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

where $\Gamma_{it\mathbf{b}}$ is the individual-specific *demand synergy* among the products in bundle $\mathbf{b}$, 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 unobserved individual-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 the market t-specific average utility for bundle $\mathbf{b}$, $\mu_{it\mathbf{b}}$ is an unobserved individual-specific 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 purchasing 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, joint purchase brings more utility. Conversely, when $\Gamma_{it\mathbf{b}} < 0$, from i's perspective the separate purchase of each $j \in \mathbf{b}$ brings more utility than their joint purchase. As we discuss below, in applications with observable bundle-specific characteristics (e.g., bundle-specific discounts), one can specify $\Gamma_{it\mathbf{b}}$ in terms of these characteristics.

We now turn to the distributional assumptions for the unobserved components of utility: $\mu_{it\mathbf{b}} = \sum_{j\in\mathbf{b}} \mu_{itj} + \zeta_{it\mathbf{b}}$ and $\varepsilon_{it\mathbf{b}}$ for each $\mathbf{b} \in \mathbf{C}_t$. We assume that $\mu_{it\mathbf{b}}$ can be specified 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 $F(\cdot;\Sigma_F)$, where Σ_F is a finite-dimensional parameter in a connected compact set $\Theta_{\Sigma_F} \subset \mathbb{R}^P$. As is typical, $\mu_{it\mathbf{b}}(\cdot)$ can also be a function of observable demographics (e.g., i's income) and/or observable market-, product-, and bundle-specific characteristics (e.g., the price of bundle $\mathbf{b}$ in market t). The error term $\varepsilon_{it\mathbf{b}}$ is assumed to be i.i.d. Gumbel.

Even though we make the assumption that $\varepsilon_{it\mathbf{b}}$ is i.i.d. Gumbel, 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 $F(\cdot;\Sigma_F)$ 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 exhibit positive demand synergies for some individuals and negative for others.

Denote the 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 bundles $\mathbf{b} \in \mathbf{C}_{t2}$, but also those of the single products $\mathbf{b} \in \mathbf{C}_{t1} \setminus \mathbf{C}_{t2}$ (where $\mathbf{C}_{t1} \setminus \mathbf{C}_{t2} = \mathbf{J}_t$): given that any single product has zero demand synergies, our notation for $\mathbf{b} = j \in \mathbf{C}_{t1} \setminus \mathbf{C}_{t2}$ is just $\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}) = \delta_{tj}$. Given our distributional assumptions, the market share function of $\mathbf{b} \in \mathbf{C}_t$ for individuals in market t takes the mixed logit form:

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

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

Complementarity and Substitutability. Following Samuelson (1974) and Gentzkow (2007), we rely on the classic notion of Hicksian complementarity: we consider two products as complements (substitutes) whenever their cross-price elasticity of demand is negative (positive).[7] In a model similar to (3) with two products, j and k, and constant demand synergy parameters $\Gamma_{it(j,k)} = \Gamma_{(j,k)}$, Gentzkow (2007) shows that j and k are complements (substitutes) whenever $\Gamma_{(j,k)} > 0$ ($\Gamma_{(j,k)} < 0$). On the one hand, with more products and heterogeneous demand synergies, the relationship between Hicksian complementarity and $\Gamma_{it\mathbf{b}}$ is less clear-cut and the topic of ongoing research (Iaria and Wang, 2019). On the other, though, standard models of demand for single products—obtained by constraining $\Gamma_{it\mathbf{b}} = -\infty$ for all i's, t's, and $\mathbf{b} \in \mathbf{C}_{t2}$—rule out the possibility of Hicksian complementarity and force any two products to be substitutes. In this paper, we take a pragmatic approach and regard the complementarity or substitutability between products as an empirical question to be answered after the estimation of model (3).

Interpretation of Demand Synergies. Model (3) is agnostic about the exact meaning of $\Gamma_{it\mathbf{b}}$, which is a catch-all parameter that can reflect, for example, synergies in consumption, shopping costs, and preference for variety. In Gentzkow (2007)'s demand for on-line and printed newspapers, $\Gamma_{it\mathbf{b}}$ captures synergies in the consumption of the different news outlets. However, demand synergies—and consequently Hicksian complementarity—can also arise, for example, because of shopping costs (Pozzi (2012) and Thomassen et al. (2017)) or preference for variety within households (Hendel (1999) and Dubé (2004)). If individuals face shopping costs every time they visit a store, they may prefer to purchase all of their products at once rather than over several trips (one-stop shoppers). Moreover, if households delegate grocery shopping to one person, then preference for variety may lead to the purchase of multiple products on any shopping trip to accommodate the different needs within the household. Our model can rationalize shopping costs with positive demand synergies and, as we show in Appendix 8.1, preference for variety with some additional structure on the demand synergy parameters.

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)

[7]In our application, we rule out income effects so that gross complementarity (in terms of elasticities of Marshallian demands) and Hicksian complementarity (in terms of elasticities of compensated demands) coincide. For discussions about complementarity in models of demand for bundles similar to those studied here, see Manzini et al. (2018), Dubé (2019), and Iaria and Wang (2019).

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. Disentangling demand synergies from these random intercepts is the key identification challenge in models of demand for bundles: as shown by Gentzkow (2007), not accounting for possible correlations across the indirect utilities of bundles with overlapping products may lead to finding spurious demand synergies and Hicksian complementarities.

Average Utilities and Price Endogeneity. We treat the average utility δ_{tj} as a fixed effect to be identified and estimated, being unspecific about its exact dependence on price and other observed or unobserved market-product specific characteristics. For example, following Berry (1994) and BLP, a classical linear specification is $\delta_{tj} = x_{tj}\tau + \alpha p_{tj} + \xi_{tj}$, where x_{tj} is a vector of exogenous observed characteristics, p_{tj} is the observed price, $(\tau, \alpha)^{\mathrm{T}}$ is a vector of preference parameters, and ξ_{tj} is a residual unobserved to the econometrician but observed to 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 $\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}) = \sum_{j\in\mathbf{b}} \delta_{tj} + \Gamma_{t\mathbf{b}}$, with $\Gamma_{t\mathbf{b}}$ constrained to be constant across markets with the same market-bundle specific observables. In particular, as detailed in Assumption 2 below, we require $\Gamma_{t\mathbf{b}} = \Gamma_{\mathbf{b}} + g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g)$, where $\Gamma_{\mathbf{b}}$ is a bundle-specific fixed effect and $g(\cdot, \cdot; \Sigma_g)$ is a function parametrized by Σ_g of the observed characteristics $x_{t\mathbf{b}}$ and of the observed price surcharge/discount $p_{t\mathbf{b}}$ (the difference between the price of bundle $\mathbf{b}$ and $\sum_{j\in\mathbf{b}} p_{tj}$). For example, one can specify $g(\cdot, \cdot; \Sigma_g)$ as $g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \tau, \alpha) = x_{t\mathbf{b}}\tau + \alpha p_{t\mathbf{b}}$. While our assumptions on $\Gamma_{t\mathbf{b}}$ allow δ_{tj} to be any arbitrary function of $(x_{tj}, p_{tj}, \xi_{tj})$, they restrict the functional form of the market-bundle specific unobservables on the basis of which firms choose prices. For instance, with the above linear specifications for δ_{tj} and $\Gamma_{t\mathbf{b}}$, we have $\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}) = (x_{t\mathbf{b}} + \sum_{j\in\mathbf{b}} x_{tj})\tau + \alpha(p_{t\mathbf{b}} + \sum_{j\in\mathbf{b}} p_{tj}) + \sum_{j\in\mathbf{b}} \xi_{tj} + \Gamma_{\mathbf{b}}$, with the market-bundle specific unobservable restricted to $\sum_{j\in\mathbf{b}} \xi_{tj} + \Gamma_{\mathbf{b}}$.

Our global identification arguments further require restrictions 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 to the econometrician but identifiable from observed market shares and prices.

3 Local Identification

Suppose that the econometrician observes without error the market shares $\mathscr{s}_{t\mathbf{b}}$ of each $\mathbf{b} \in \mathbf{C}_{t1}$ for each independent market $t = 1, ..., T$.[8,9] 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. Similar to Berry and Haile (2014), our notion of identification concerns the conditions under which

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

has a unique solution for $t \in \mathbf{T}$ 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 $s_{t\mathbf{b}}(\delta'_t(\Gamma'_t); \Sigma'_F)$ is defined in (3). 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 $s_t(\cdot; \Sigma'_F) = \big(s_{t\mathbf{b}}(\cdot; \Sigma'_F)\big)_{\mathbf{b} \in \mathbf{C}_{t1}}$ and $\mathscr{s}_t = (\mathscr{s}_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_{t1}}$.

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

Definition 1 constrains our discussion of identification to the existence of a unique solution to system (4) in mixed logit model (3). We will refer to the existence of multiple solutions to this specific problem as to lack of identification. 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 et al. (2013), our identification arguments rely on demand inverses derived from (4). Define the inverse market share for market $t \in \mathbf{T}$ as:

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

where $s_{t\mathbf{b}}^{-1}(\cdot; \Sigma_F)$ is the inverse market share for market $t = 1, ..., T$ and $\mathbf{b} \in \mathbf{C}_{t1}$, 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 for market $t \in \mathbf{T}$. The next Assumption imposes some regularity conditions on the parametric distribution of the random coefficients (first requirement) and that the products belonging to any bundle can also be purchased individually (second requirement).

[8]This is only for the purpose of identification, in estimation we consider the case of observed market shares subject to sampling error.

[9]Sher and Kim (2014), Allen and Rehbeck (2019), and Wang (2019) study a different identification problem, where only the product-level market shares, rather than the bundle-level market shares, are observed.

Assumption 1.

1. *The density of* β_{it}, $\frac{dF(\beta_{it};\Sigma'_F)}{d\beta_{it}}$, *is continuously differentiable with respect to* Σ'_F *for any* β_{it}.

2. *If* $\mathbf{b} \in \mathbf{C}_{t2}$, *then* $j \in \mathbf{J}_t$ *for any* $j \in \mathbf{b}$.

The next Lemma verifies the sufficient conditions by Berry et al. (2013) for the bundle-level demand inverse (5) to be a continuously differentiable function.

Lemma 1.

- *For any given* $\Sigma'_F \in \Theta_{\Sigma_F}$, *the inverse market share (5) is a function: for each* $\mathfrak{z}_t \in \mathbf{S}_{t1}$, *there exists a unique* $\delta'_t \in \mathbb{R}^{C_{t1}}$ *such that* $s_t(\delta'_t;\Sigma'_F) = \mathfrak{z}_t$.

- *Given Assumption 1, the inverse market share,* $s_t^{-1}(\mathfrak{z}'_t;\Sigma'_F)$, *is continuously differentiable with respect to* $(\mathfrak{z}'_t,\Sigma'_F)$ *in a neighbourhood of* $(\mathfrak{z}_t,\Sigma_F)$.

Proof. See Appendix 8.2. □

In the online supplement, we illustrate that even simple versions of model (3) raise identification concerns. 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 multiple markets will help identification. In what follows, we study identification under this restriction.

Assumption 2. $\Gamma_{t\mathbf{b}} = \Gamma_{\mathbf{b}} + g(x_{t\mathbf{b}}, p_{t\mathbf{b}};\Sigma_g)$ *for* $t \in \mathbf{T}$ *and* $\mathbf{b} \in \mathbf{C}_{t2}$, *where* $\Gamma_{\mathbf{b}}$ *is a bundle-specific fixed effect,* $x_{t\mathbf{b}}$ *a vector of observed market-bundle specific characteristics,* $p_{t\mathbf{b}}$ *an observed price surcharge/discount for the joint purchase of the products in the bundle, and* $g(\cdot,\cdot;\Sigma_g)$ *a function of* $(x_{t\mathbf{b}}, p_{t\mathbf{b}})$ *known up to and continuously differentiable with respect to* $\Sigma_g \in \Theta_{\Sigma_g} \subseteq \mathbb{R}^D$.

Assumption 2 restricts the variation in $\Gamma_{t\mathbf{b}}$ across markets to be fully captured by the variation in the observables $(x_{t\mathbf{b}}, p_{t\mathbf{b}})$ through the parametric function $g(x_{t\mathbf{b}}, p_{t\mathbf{b}};\Sigma_g)$. This allows to reduce the dimensionality of the collection of average demand synergies from $\sum_{t=1}^T C_{t2}$ to $D + C_2$ and in particular to treat $\Gamma_{\mathbf{b}}$ as a bundle-specific fixed effect to be identified and estimated. Note that, even though Assumption 2 requires all markets with given $(x_{t\mathbf{b}}, p_{t\mathbf{b}})$ to have the same average demand synergy $\Gamma_{\mathbf{b}} + g(x_{t\mathbf{b}}, p_{t\mathbf{b}};\Sigma_g)$, each individual in each market is allowed to have a specific demand synergy deviation $\zeta_{it\mathbf{b}}$, so that $\Gamma_{it\mathbf{b}}$ may potentially differ across individuals for any given market and bundle. Moreover, as we illustrate below, in applications with a large number of markets with overlapping choice sets, Assumption 2 potentially leads to many over-identifying moment restrictions and can be weakened, so to allow for more flexible specifications of $\Gamma_{t\mathbf{b}}$. In those cases, one could for example specify $(\Gamma_{t\mathbf{b}}, \Sigma_{tg}) = (\Gamma_{1\mathbf{b}}, \Sigma_{1g})$ for $t = 1, ..., T_1$, $(\Gamma_{t\mathbf{b}}, \Sigma_{tg}) = (\Gamma_{2\mathbf{b}}, \Sigma_{2g})$ for $t = T_1 + 1, ..., T_2$, and so on until each t belonged to one of Q groups of "similar" markets with $(\Gamma_{1\mathbf{b}}, \Sigma_{1g}) \neq (\Gamma_{2\mathbf{b}}, \Sigma_{2g}) \neq ... \neq (\Gamma_{Q\mathbf{b}}, \Sigma_{Qg})$.

Remark 1. *Assumption 2 gives rise to testable implications and can be verified in practice. In Appendix 8.7, we present a specification test that builds on partial identification methods. Essentially, the proposed test checks whether there exists at least one profile of parameters* $(\delta_{1\mathbf{J}_1}, ..., \delta_{T\mathbf{J}_T}, \Gamma_1, ..., \Gamma_T, \Sigma_F)$ *satisfying Assumption 2 that solves demand system (4). A rejection of the test is evidence against Assumption 2 and highlights its incoherence with the data.*

The presence of random coefficients, i.e. $\dim(\Sigma_F) > 0$, 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 including additional instruments beyond those necessary to address price endogeneity. However, in the case of Gentzkow (2007)'s demand for bundles, the specific structure that links the indirect utilities of bundles to those of single products 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 to those necessary to address price endogeneity.

Due to Lemma 1 and Assumption 2, at the true parameters Σ_F and market shares $\mathscr{s}_t$, one can re-express the first line of system (4) as:

$$\begin{aligned} &\sum_{j\in\mathbf{b}} \delta_{tj} + \Gamma_{\mathbf{b}} + g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g) = s_{t\mathbf{b}}^{-1}(\mathscr{s}_t; \Sigma_F), \text{ for bundle } \mathbf{b} \in \mathbf{C}_{t2} \\ &\delta_{tj} = s_{tj}^{-1}(\mathscr{s}_t; \Sigma_F), \text{ for product } j \in \mathbf{b}. \end{aligned} \tag{6}$$

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

$$\Gamma_{\mathbf{b}} = s_{t\mathbf{b}}^{-1}(\mathscr{s}_t; \Sigma_F) - \sum_{j\in\mathbf{b}} s_{tj}^{-1}(\mathscr{s}_t; \Sigma_F) - g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g), \tag{7}$$

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

$$s_{t\mathbf{b}}^{-1}(\mathscr{s}_t; \Sigma_F) - \sum_{j\in\mathbf{b}} s_{tj}^{-1}(\mathscr{s}_t; \Sigma_F) - g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g) = s_{t'\mathbf{b}}^{-1}(\mathscr{s}_{t'}; \Sigma_F) - \sum_{j\in\mathbf{b}} s_{t'j}^{-1}(\mathscr{s}_{t'}; \Sigma_F) - g(x_{t'\mathbf{b}}, p_{t'\mathbf{b}}; \Sigma_g). \tag{8}$$

Our identification strategy exploits 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 below, under certain conditions, these moment restrictions can uniquely determine the true parameters $\Sigma = (\Sigma_F, \Sigma_g)$. Then, due to (7), the true parameters $\Sigma = (\Sigma_F, \Sigma_g)$ can uniquely determine the remaining portion $\Gamma_{\mathbf{b}}$ of the true demand synergies, for any $\mathbf{b} \in \mathbf{C}_{t2}$. Denote $g_t(\Sigma_g) = (g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g))_{\mathbf{b}\in\mathbf{C}_{t2}}$. Finally, because of Lemma 1, one can uniquely recover $\delta_t(\Gamma + g_t(\Sigma_g)) = (\delta_{t1}, ..., \delta_{tJ_t}, (\delta_{t\mathbf{b}}(\Gamma_{\mathbf{b}} + g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g)))_{\mathbf{b}\in\mathbf{C}_{t2}})$.

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 $\mathbf{T_b} = \{t : \mathbf{b} \in \mathbf{C}_{t2}, t \in \mathbf{T}\}$. If $\mathbf{T_b}$ has more than one element, we order them from t_1 to $t_{|\mathbf{T_b}|}$. 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 then obtain $|\mathbf{T_b}| - 1$ moment conditions:[10]

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

Moment conditions (9) rely on relationship (7) and the fact that markets t_a and t_{a+1} have the same bundle-specific fixed effect $\Gamma_{\mathbf{b}}$. As a consequence, 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 $\Sigma' = (\Sigma'_F, \Sigma'_g) \in \Theta_\Sigma = \Theta_{\Sigma_F} \times \Theta_{\Sigma_g}$ that stacks together the above moment conditions for all the bundles 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 $P + D = \dim(\Sigma')$ unknowns.

In what follows, inspired by Rothenberg (1971), we show that full column rank condition $\text{rank}(\frac{\partial m(\Sigma')}{\partial \Sigma'})\big|_{\Sigma'=\Sigma} = \dim(\Sigma) = P + D$ is necessary and sufficient for identification among the *rank regular* $\Sigma \in \Theta_\Sigma$.[11,12] 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, then Σ is rank regular.[13]

Theorem 1. ***Local Identification:*** *Suppose 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 8.5 □

[10]For notational simplicity, we suppress the dependence of the moment conditions from the market-bundle specific observables $(x_{t\mathbf{b}}, p_{t\mathbf{b}})_{t=1}^T$.

[11]$\Sigma \in \Theta_\Sigma$ 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 $\Sigma' \in U$.

[12]Rothenberg (1971) shows the usefulness of the concept of rank regularity for local identification in non-linear models. Our Theorem 1 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 8.3.

[13]In fact, $\left[\frac{\partial m(\Sigma')}{\partial \Sigma'}\right]^{\mathrm{T}} \left[\frac{\partial m(\Sigma')}{\partial \Sigma'}\right]$ has positive determinant at $\Sigma' = \Sigma$. Moreover, $\frac{\partial m(\Sigma')}{\partial \Sigma'}$ is continuously differentiable with respect to Σ'. Then, the determinant of $\left[\frac{\partial m(\Sigma')}{\partial \Sigma'}\right]^{\mathrm{T}} \left[\frac{\partial m(\Sigma')}{\partial \Sigma'}\right]$ 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 $P + D$ in the same neighbourhood of $\Sigma' = \Sigma$.

Theorem 1 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 1 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. Specifically, suppose that Σ is rank regular and that its dimension, $P + D$, 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 the number of its columns, $P + D$. As a consequence, $\frac{\partial m(\Sigma')}{\partial \Sigma'}\big|_{\Sigma'=\Sigma}$ is not of full column rank and model (3) is not identified.

While theoretically useful, the concept of rank regularity is abstract and not easily verifiable. 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.[14]

Corollary 1. ***Lack of Local Identification:*** *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, $P + D$. *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 8.6. □

4 Global Identification

Up to this point, we have 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. 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 motivated by estimation convenience: as detailed in section 5, our global identification conditions imply a convenient MLE.

Denote by $m(\Sigma'; \mathbf{T}_0)$ moment conditions (9) constructed from the subset of markets $\mathbf{T}_0 \subsetneq \mathbf{T}$ and evaluated at $\Sigma' = (\Sigma'_F, \Sigma'_g)$. The starting point of our global identification argument is to restrict the number of solutions to system (9) with the following testable Assumption:

[14]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.

Assumption 3. *There exists* $\mathbf{T}_0 \subsetneq \mathbf{T}$ *such that* $m(\Sigma'; \mathbf{T}_0) = 0$ *has a finite number of solutions in* Θ_Σ.

Assumption 3 alleviates the severity of the identification problem to a finite collection of observationally equivalent candidates. All of our global identification results go through also in the more general case of a countable collection of solutions. However, for notational simplicity, we limit our exposition to the finite case. Assumption 3 implies local identification, but is weaker than the typical rank conditions used to achieve global identification. For example, the classic results by Rothenberg (1971) require the Jacobian of the gradient of the log-likelihood function to be non-singular everywhere, so that the log-likelihood function is strictly concave (Bowden, 1973). While strict concavity is guaranteed by logit and probit models (Amemiya (1985) pp. 273-274), it is not by mixed logit models. Coherently with mixed logit model (3), Assumption 3 does not impose strict concavity of the log-likelihood function.

Remark 2. *While Assumption 3 is high-level, in Appendix 8.7 we present a verifiable sufficient condition that implies it (Proposition 2):* $\frac{\partial m(\Sigma'; \mathbf{T}_0)}{\partial \Sigma'}$ *is of full column rank when evaluated at any of the solutions to* $m(\Sigma'; \mathbf{T}_0) = 0$. *Building on the partial identification methods by Chernozhukov et al. (2007) and by Romano and Shaikh (2012), in Appendix 8.7 we also propose a testing procedure to verify this sufficient condition in practice. A rejection of the test is evidence in support of Assumption 3.*

Assumption 3 is not new to the global identification literature and is also used, for example, by Komunjer (2012). To obtain global identification, Komunjer (2012) additionally requires the moment function to have non-negative Jacobian and to be proper.[15] We avoid these further restrictions by relying on the following real analytic property of the mixed logit model.[16]

Theorem 2. ***Real Analytic Property:*** *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 8.8. □

Theorem 2 shows the market share function of the mixed logit model to be real analytic with respect to the average utilities given any distribution of random coefficients (parametric or non-parametric). Fox et al. (2012) and il Kim (2014) also exploit the real analytic properties of logit models to achieve global identification, but in more restrictive frameworks. il Kim (2014) shows the real analyticity of multinomial logit and nested logit models (section IV), while Fox et al. (2012) show it for mixed logit models with random coefficients defined over

[15] A function $f : X \to Y$ between two topological spaces is *proper* if the preimage of every compact set in Y is compact in X.

[16] 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 $x \in U$.

compact supports (Lemma 5 and section 6)—thereby ruling out, for example, the normal and log-normal distributions.

While our local identification results do not rely on the nature of the variation in $\delta_{t\mathbf{J}_t}$, our global identification depends on whether the variation in $\delta_{t\mathbf{J}_t}$ is exogenous across markets: price endogeneity restricts this variation and leads to additional difficulties. 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 exogenous variation in $\delta_{t\mathbf{J}_t}$, and that 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 Assumption 3, denote the finite set of solutions to $m(\Sigma'; \mathbf{T}_0) = 0$ in Θ_Σ by $S = \{\Sigma^r : r = 0, ..., R\}$, where $\Sigma^0 = (\Sigma_F^0, \Sigma_g^0)$ represents the true value $\Sigma = (\Sigma_F, \Sigma_g)$. On the basis of Lemma 1, define the corresponding Γ^r for $r = 0, 1, ..., R$. The real analytic property of $s_t(\delta_t'; \Sigma_F')$ allows 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 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 number of zero measure sets has still zero measure, the real analytic property—combined with Assumption 3—ensures global identification almost everywhere given the additional variation provided by $\delta_{t\mathbf{J}_t}$, $t \in \mathbf{T} \setminus \mathbf{T}_0$.

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} is the number of bundles and C_{t1} the number of inside options (i.e., bundles plus single products). 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 single products and the rows bundles. Each row of M_i^1 identifies with -1's the product composition of the corresponding bundle. M_t^2 is instead an identity matrix $\mathbf{I}$ of dimension $C_{t2} \times C_{t2}$, with the rows corresponding to bundles. For example, suppose the choice set (without outside option) 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,

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

Remember that $g_t(\Sigma_g) = (g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g))_{\mathbf{b}\in\mathbf{C}_{t2}}$ and that $\delta_t(\Gamma + g_t(\Sigma_g)) = (\delta_{t\mathbf{J}_t}, (\delta_{t\mathbf{b}}(\Gamma_\mathbf{b} + g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g)))_{\mathbf{b}\in\mathbf{C}_{t2}})$, where $\delta_{t\mathbf{b}}(\Gamma_\mathbf{b} + g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g)) = \sum_{j\in\mathbf{b}} \delta_{tj} + \Gamma_\mathbf{b} + g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g)$. For $r = 1, ..., R$, define:

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

Denote by $\delta^0_{t\mathbf{J}_t}$ the true value of $\delta_{t\mathbf{J}_t}$ for which $s_t(\delta_t(\Gamma^0 + g_t(\Sigma^0_g)); \Sigma^0_F) = \mathcal{s}_t$ and define $\Delta^{\mathrm{ID}} = \cap_{r=1,\ldots,R} \Delta^{\mathrm{ID}}_r$.

Theorem 3. ***Global Identification with Exogenous Prices:*** *Suppose Assumptions 1–3 hold and Θ_Σ is compact. Then, it follows that:*

- *System (9) has a unique solution in Θ_Σ and model (3) is globally identified if and only if $(\delta^0_{t\mathbf{J}_t})_{t\in\mathbf{T}\setminus\mathbf{T}_0} \in \Delta^{ID}$.*
- *If $\Delta^{ID}_r \neq \emptyset$ for $r = 1, ..., R$, then 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 8.9. □

While the first result of Theorem 3 provides necessary and 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

We now extend the global identification results from Theorem 3 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. We add mild restrictions to the price-setting model and 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) for the 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 argument is different and does not require the instrumental variables (i.e., the cost shifters) to be observed to the econometrician but only to be identifiable.

Similar to BLP, we specify the average utility δ_{tj} as additively separable in a systematic component and an unobserved residual: $\delta_{tj} = \Delta_{tj}(x_{tj}, p_{tj}) + \xi_{tj}$, with x_{tj} a vector of observed exogenous characteristics, p_{tj} the observed endogenous price, $\Delta_{tj}(\cdot, \cdot)$ any arbitrary function of (x_{tj}, p_{tj}) (potentially different across markets and products), and ξ_{tj} a residual unobserved to the econometrician. Even though we rely on the additive separability of $\Delta_{tj}(x_{tj}, p_{tj})$ and ξ_{tj}, the target of our identification is still their sum δ_{tj}. Endogeneity arises whenever firms choose prices (also) on the basis of the market-specific residuals, which we denote by $\xi_{t\mathbf{J}_t} = (\xi_{tj})_{j\in\mathbf{J}_t} \in \mathbb{R}^{J_t}$. Because we essentially treat each δ_{tj} as a fixed effect, price endogeneity complicates global identification to the extent that it constrains the variation of $\delta_{t\mathbf{J}_t}$ across markets (the key identifying variation used in Theorem 3). As an extreme example, suppose that prices are chosen so that $\Delta_{tj}(x_{tj}, p_{tj}) = -\xi_{tj}$, then $\delta_{tj} = 0$ for every t and j. This rules out *any* variability in $\delta_{t\mathbf{J}_t}$, introducing the need for alternative sources of identification. To simplify exposition, in what follows we sometimes drop the dependence on $\Delta_{tj}(x_{tj}, p_{tj})$ from our notation.

Here we discuss the case of pure components pricing, where each firm chooses the prices of the individual products it owns and the price of any bundle is given by the sum of the prices of its components. With pure components pricing, the econometrician observes the prices of the individual products $p_{t\mathbf{J}_t} = (p_{tj})_{j \in \mathbf{J}_t}$, while the price surcharges/discounts for the joint purchase of products in bundles are all constrained to zero, so that $p_{t\mathbf{b}} = 0$ and $g_{t\mathbf{b}}(\Sigma'_g) = g_{t\mathbf{b}}(x_{t\mathbf{b}}, 0; \Sigma'_g)$ for $t = 1, ..., T$ and $\mathbf{b} \in \mathbf{C}_{t2}$. As we illustrate in the online supplement, our arguments can be readily modified to accommodate alternative pricing strategies such as mixed bundling (see Armstrong (2016a) for a survey on non-linear pricing).[17]

Denote 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. These cost shifters could for example be the marginal costs of the products sold in market t. Similar to $\xi_{t\mathbf{J}_t}$, also the cost shifters $c_{t\mathbf{J}_t}$ are assumed to be unobserved to the econometrician. In this sense, cost shifters can be seen as "unobserved" instruments: their existence provides exogenous identifying variation, but they do not need to be observed to the econometrician. As for the case of exogenous average utilities, we propose a characterization of the set of unobservables $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$ that suffices for the global identification of $(\delta_{t\mathbf{J}_t}, \Gamma, \Sigma)$.

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 $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}) \in D_{t\xi} \times D_{tc}$. Denote 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 $p_{t\mathbf{J}_t} \in p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$.

Assumption 4.

- *(Cost Shifters at the Product-Level)* 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* $(\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t}) \in \{(\xi'_{t\mathbf{J}_t}, p'_{t\mathbf{J}_t}) : \xi'_{t\mathbf{J}_t} \in D_{t\xi}, p'_{t\mathbf{J}_t} \in \mathbf{P}_t(\xi_{t\mathbf{J}_t})\}$*:* $c_{t\mathbf{J}_t} = \phi_t(\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t})$.

The second part of Assumption 4 resembles Assumption 7b by Berry and Haile (2014) and is the key to our global identification with price 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 (Remark 1, p. 1766). Their result follows from the assumption of "connected substitutes" on the demand system (Definition 1, p. 1759): loosely speaking, this rules out any negative cross-price elasticity between any two products. In the case of

[17] With mixed bundling pricing, every 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. In this case, the price surcharge/discount for the joint purchase of products in bundles, $p_{t\mathbf{b}}$, may differ from zero for any t and $\mathbf{b} \in \mathbf{C}_{t2}$.

pure components pricing, the relevant demand system has only J_t product-level equations (the system of product-level market shares) rather than C_{t1} bundle-level equations. While model (3) satisfies the connected substitutes property at the bundle-level, it may not at the product-level (i.e., products may be complements) and hence Remark 1 by Berry and Haile (2014) does not apply to our case.

By combining the bundle-level connected substitutes property with the specific utility structure of model (3) under Assumption 2, in Appendix 8.10 we show that Assumption 4 is satisfied by common pure components pricing models. We show that it 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 (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 infinitely many equilibria.

Denote by $s_t(\delta'_t(\Gamma' + g_t(\Sigma'_g)); p'_{t\mathbf{J}_t}, \Sigma'_F)$ the market share function in market t evaluated at prices $p'_{t\mathbf{J}_t} = (p'_{tj})_{j\in\mathbf{J}_t}$ and structural parameters $(\delta'_{t\mathbf{J}_t}, \Gamma', \Sigma')$, and remember that $\delta'_{tj} = \Delta'_{tj}(x_{tj}, p_{tj}) + \xi'_{tj}$ for each $j \in \mathbf{J}_t$. Given Assumption 3, define for each $r = 1, ..., R$:

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

and $\Xi^{\mathrm{ID}} = \cap_{r=1}^R \Xi_r^{\mathrm{ID}}$. We make the following technical Assumption:

Assumption 5. *For every $r = 1, ..., R$, there exists $t \in \mathbf{T}\setminus\mathbf{T}_0$, so that for almost every $p_{t\mathbf{J}_t} \in \mathbf{P}_t$, there exists $\xi'_{t\mathbf{J}_t}$, such that $\Gamma^r + g_t(\Sigma_g^r) \neq M_t s_t^{-1}(s_t(\delta'_t(\Gamma^0 + g_t(\Sigma_g^0)); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r)$.*

Even though Assumption 5 is abstract, it is implied by more concrete conditions. The following Corollary shows that, for example, by strengthening the real analytic property of mixed logit models from Theorem 2, Assumption 5 is satisfied:

Corollary 2. *Suppose that the following conditions hold:*

- *For $r = 1, ..., R$, $\Xi_r^{ID} \neq \emptyset$.*
- *For any $t \in \mathbf{T}\setminus\mathbf{T}_0$ and (Γ, Σ), $s_t(\delta'_t(\Gamma + g_t(\Sigma_g)); p'_{t\mathbf{J}_t}, \Sigma_F)$ is real analytic with respect to $(\delta'_{t\mathbf{J}_t}, p'_{t\mathbf{J}_t})$.*

Then, Assumption 5 holds.

Proof. See Appendix 8.11. □

Corollary 2 tightens the real analyticity of the market share function to hold also with respect to the prices $p'_{t\mathbf{J}_t}$ (in addition to the average utilities $\delta'_{t\mathbf{J}_t}$). If price enters the indirect utility linearly (as is typical in applied work), then Corollary 2 will hold when the price coefficient is for example constant, or bounded, or when its moments increase at most exponentially.

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 $s_t(\delta_t(\Gamma^0+g_t(\Sigma^0_g)); p^0_{t\mathbf{J}_t}, \Sigma^0_F) = \mathcal{s}_t$. We now present the main identification result of the paper.

Theorem 4. ***Global Identification with Endogenous Prices:*** *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 Assumption 5 holds, 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 8.12. □

As for Theorem 3, the first part of Theorem 4 provides sufficient conditions for global identification, while the second highlights that global identification will be achieved almost everywhere.

5 Estimation

We propose a 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 et al. (2015), Grzybowski and Verboven (2016), Ruiz et al. (2017), and Ershov et al. (2018)). We consider asymptotics over the number of individuals I within each market, keeping fixed the number of markets and bundles, and demonstrate that our identification conditions imply the proposed MLE to be consistent and asymptotically normal.

Even though theoretically attractive, the standard MLE of model (3) is subject to a challenge of dimensionality even under Assumption 2: the number of demand parameters can still be too large to be handled numerically (Berry et al., 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 Σ_F for the distribution of random coefficients, D parameters Σ_g for the function g_t, and $J \times T$ average utility parameters $(\delta_{1\mathbf{J}_1}, ..., \delta_{T\mathbf{J}_T})$. The estimation of $J^K + P + D + J \times T$ parameters may be hard, especially because identification requires a large T.

We propose to reduce the dimensionality of the MLE's numerical search by means of a novel demand inversion specific to Gentzkow (2007)'s model that concentrates $(\delta'_{1\mathbf{J}_1}, ..., \delta'_{T\mathbf{J}_T})$ out of the likelihood function.[18] As a consequence, our proposed MLE effectively reduces the numerical search from $(\delta'_{1\mathbf{J}_1}, ..., \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma')$, i.e. $J^K + P + D + J \times T$ parameters, to (Γ', Σ'), i.e. $J^K + P + D$ parameters.

[18] As we clarify below, our demand inverse differs from the classic one by Berry (1994) and Berry et al. (2013), which in our context corresponds to the demand inverse presented in Lemma 1.

Remark 3. *Our estimation discussion focuses on the case of exogenous average utilities. However, when the assumptions from the previous section hold and the model is globally identified, the estimation results presented below will also hold for the case of price endogeneity with no modification. The exogenous cost shifters that play the role of instruments in our identification arguments need to exist but do not need to be observed. The estimation of* $(\delta'_{1\mathbf{J}_1}, ..., \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma')$ *will not require the "explicit" use of instruments also in the presence of price endogeneity.*

5.1 Invertibility of Product-Level Market Shares

Here we propose a novel demand inverse designed to handle complementarity among products in models along the lines of Gentzkow (2007)'s. For any given value of the other parameters, we establish a one-to-one mapping between the observed product-level market shares and the market-product specific average utilities. We then illustrate how this demand inverse can be used to greatly simplify the practical implementation of the MLE of demand for bundles.

Define the observed product-level market share of product $j \in \mathbf{J}_t$ as $\mathscr{s}_{tj.} = \sum_{\mathbf{b} \in \mathbf{C}_{t1}: j \in \mathbf{b}} \mathscr{s}_{t\mathbf{b}}$ and denote the vector stacking $\mathscr{s}_{tj.}$ for all products in market t by $\mathscr{s}_{t\mathbf{J}_t.} = (\mathscr{s}_{tj.})_{j \in \mathbf{J}_t}$. Similarly, define the product-level market share function of each product $j \in \mathbf{J}_t$ as $s_{tj.}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma') = \sum_{\mathbf{b} \in \mathbf{C}_{t1}: j \in \mathbf{b}} s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F)$ and denote the vector stacking $s_{tj.}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma')$ for all products in market t by $s_{t\mathbf{J}_t.}(\cdot; \Gamma', \Sigma') = (s_{tj.}(\cdot; \Gamma', \Sigma'))_{j \in \mathbf{J}_t}$.

Theorem 5. ***Demand Inverse:*** *Suppose that Assumptions 1 and 2 hold. Then, for any* $(\Gamma', \Sigma') \in \Theta_\Gamma \times \Theta_\Sigma$*, there exists at most one* $\delta'_{t\mathbf{J}_t}$ *such that* $s_{t\mathbf{J}_t.}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma') = \mathscr{s}_{t\mathbf{J}_t.}$.

Proof. See Appendix 8.13. □

When (Γ', Σ') is equal to the true value (Γ, Σ), Theorem 5 implies that the only $\delta'_{t\mathbf{J}_t}$ that satisfies $s_{t\mathbf{J}_t.}(\delta'_{t\mathbf{J}_t}; \Gamma, \Sigma) = \mathscr{s}_{t\mathbf{J}_t.}$ is the true $\delta_{t\mathbf{J}_t}$. As a result, the function $s_{t\mathbf{J}_t.}(\cdot; \Gamma, \Sigma)$ is globally invertible. When $(\Gamma', \Sigma') \neq (\Gamma, \Sigma)$, it is possible that there is no $\delta'_{t\mathbf{J}_t}$ such that $s_{t\mathbf{J}_t.}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma') = \mathscr{s}_{t\mathbf{J}_t.}$.[19] Because the existence of some $\delta'_{t\mathbf{J}_t}$ that rationalizes the observed product-level market shares can always be verified numerically (following the procedure outlined below), in what follows we assume it and denote the bijection mapping $\mathscr{s}_{t\mathbf{J}_t.}$ to $\delta'_{t\mathbf{J}_t}$ by:

$$\begin{aligned} \delta'_{t\mathbf{J}_t} &= s^{-1}_{t\mathbf{J}_t.}(\mathscr{s}_{t\mathbf{J}_t.}; \Gamma', \Sigma') \\ &= \delta_{t\mathbf{J}_t.}(\mathscr{s}_{t\mathbf{J}_t.}; \Gamma', \Sigma'). \end{aligned}$$

Theorem 5 differs from the classic demand inverse by Berry (1994) (then generalized by Berry et al. (2013)). In our context, Berry (1994) implies a bijection between the observed *bundle*-level

[19]For example, a model of demand for single products (i.e., $\Gamma' = -\infty$) cannot rationalize situations in which the sum of the observed product-level market shares is larger than one. (This can happen because the same $\mathscr{s}_{t\mathbf{b}}$ contributes to the product-level market share of any $j \in \mathbf{b}$, giving rise to "multiple counting" of $\mathscr{s}_{t\mathbf{b}}$ when summing $\mathscr{s}_{tj.}$ over j.) In such cases, the demand inverse is therefore not feasible when evaluated at $\Gamma' = -\infty$.

market shares and the *full* vector of market-bundle specific average utilities. We rely on this classic demand inverse throughout the paper and, for completeness, adapt it to our framework in Lemma 1. Differently, Theorem 5 establishes a bijection between a *transformation* of the observed bundle-level market shares—the *product*-level market shares—and a *sub*-vector of the market-bundle specific average utilities—the market-*product* specific average utilities. While the invertibility of the product-level market shares on the basis of Berry (1994) would require the products to be substitutes, Theorem 5 applies also to the case of complementary products.

5.2 A Maximum Likelihood Estimator

We now allow for the possibility that observed market shares are subject to sampling error, due for example to the necessity of measuring them from household-level purchase data. Denote by $I_{t\mathbf{b}}$ the number of individuals in market t observed to choose $\mathbf{b}$ and by $\hat{\jmath}_{t\mathbf{b}} = \frac{I_{t\mathbf{b}}}{I}$ the corresponding observed market share. To simplify exposition, in what follows we drop any notational dependence from the observables and denote $g_t(\Sigma'_g) = (g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma'_g))_{\mathbf{b}\in\mathbf{C}_{t2}}$. The log-likelihood function evaluated at $(\delta'_{1\mathbf{J}_1}, ..., \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma')$ can be written as:

$$\ell_I(\delta'_{1\mathbf{J}_1}, ..., \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma'; \hat{\jmath}_1, ..., \hat{\jmath}_T) = \sum_{t=1}^{T} \sum_{\mathbf{b}\in\mathbf{C}_t} \hat{\jmath}_{t\mathbf{b}} \log s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F), \tag{10}$$

where $\hat{\jmath}_t = (\hat{\jmath}_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_t}$ for $t = 1, ..., T$. Denote the domain of the parameters by $\Theta = \Theta_\delta \times \Theta_\Gamma \times \Theta_\Sigma$, where Θ_δ, Θ_Γ, and Θ_Σ are compact. Given Theorem 5, we propose the following MLE that concentrates $(\delta_{1\mathbf{J}_1}, ..., \delta_{T\mathbf{J}_T})$ out of the log-likelihood function:

$$\begin{aligned}
(\hat{\Gamma}, \hat{\Sigma}) &\equiv \text{argmax}_{(\Gamma',\Sigma')\in\Theta_\Gamma\times\Theta_\Sigma} \ell_I((\delta_{t\mathbf{J}_t\cdot}(\hat{\jmath}_{t\mathbf{J}_t\cdot}; \Gamma', \Sigma'))_{t=1}^T, \Gamma', \Sigma'; \hat{\jmath}_1, ..., \hat{\jmath}_T), \\
&= \text{argmax}_{(\Gamma',\Sigma')\in\Theta_\Gamma\times\Theta_\Sigma} \ell^c_I(\Gamma', \Sigma'; \hat{\jmath}_1, ..., \hat{\jmath}_T) \\
\hat{\delta}_{t\mathbf{J}_t} &\equiv \delta_{t\mathbf{J}_t\cdot}(\hat{\jmath}_{t\mathbf{J}_t\cdot}; \hat{\Gamma}, \hat{\Sigma}), \ t = 1, ..., T.
\end{aligned} \tag{11}$$

To simplify notation, denote the true parameters $(\delta_{1\mathbf{J}_1}, ..., \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$ by $\theta = (\theta_\delta, \Gamma, \Sigma)$ and the MLE $(\hat{\theta}_\delta, \hat{\Gamma}, \hat{\Sigma})$ by $\hat{\theta}$. The next Theorem establishes the asymptotic properties of $\hat{\theta}$.

Theorem 6. ***MLE estimator:*** *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 8.14 hold. Then:*

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

Proof. See Appendix 8.14. □

Estimator (11) is neither a standard MLE nor a concentrated MLE. A standard MLE would maximize (10) with respect to (δ, Γ, Σ), while (11) only maximizes it with respect to (Γ, Σ). Differently from a concentrated MLE, which also would maximize (10) only with respect to (Γ, Σ), estimator (11) is however not as efficient as the standard MLE. The demand inverse from Theorem 5 only uses observed product-level market shares (rather than bundle-level), and this causes a loss of information in the process of concentrating out θ_δ from the log-likelihood function. MLE (11) trades-off computational ease against estimation efficiency.

Implementation. In the spirit of BLP, the demand inverse from Theorem 5 enables to break down the numerical search for $(\delta_{1\mathbf{J}_1}, ..., \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$ into two steps that can be solved sequentially while implementing (11):

Step 1. For any given guess of (Γ', Σ') and market $t = 1, ..., T$, compute $\delta'_{t\mathbf{J}_t} = \delta_{t\mathbf{J}_t\cdot}(\hat{s}_{t\mathbf{J}_t\cdot}; \Gamma', \Sigma')$ by the Newton-Raphson method as the unique solution to system $s_{t\mathbf{J}_t\cdot}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma') = \hat{s}_{t\mathbf{J}_t\cdot}$. To implement the Newton-Raphson method, note that the derivative $\frac{\partial s_{t\mathbf{J}_t\cdot}(\delta'_{t\mathbf{J}_t};\Gamma',\Sigma')}{\partial \delta'_{t\mathbf{J}_t}} = [\mathbf{I} \ \text{-}M_t^{1\mathrm{T}}]\frac{\partial s_{t\mathbf{C}_{t1}}}{\partial \delta'_{t\mathbf{C}_{t1}}}[\mathbf{I} \ \text{-}M_t^{1\mathrm{T}}]^{\mathrm{T}}$ is everywhere symmetric and positive-definite, where M_t^1 is defined prior to Theorem 3. Because the solution to the system is guaranteed to be at most unique, whenever the algorithm finds one, the numerical search can end.[20] Given this solution, compute the derivative $\frac{\partial \delta_{t\mathbf{J}_t\cdot}}{\partial(\Gamma',\Sigma')} = -\left[\frac{\partial s_{t\mathbf{J}_t\cdot}}{\partial \delta'_{t\mathbf{J}_t}}\right]^{-1}\left[\frac{\partial s_{t\mathbf{J}_t\cdot}}{\partial(\Gamma',\Sigma')}\right]$ and move on to Step 2. In case the algorithm cannot find a solution, then Theorem 5 implies that $(\Gamma', \Sigma') \neq (\Gamma, \Sigma)$: try a new guess of (Γ', Σ') and go back to the beginning of Step 1.

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

6 Empirical Illustration

We illustrate our methods in the context of the ready-to-eat (RTE) cereal industry in the USA. We revisit the classic studies by Nevo (2000, 2001), and allow for Hicksian complementarity among different brands in demand estimation. The households in our data are observed to purchase two or more different RTE cereal brands in approximately 20% of their shopping trips. In the data, we observe purchases rather than consumption. In terms of purchases, demand for bundles can arise for reasons different from synergies in consumption (as in Gentzkow (2007)):

[20] For a useful discussion about the Newton-Raphson method in the context of demand estimation, see Conlon and Gortmaker (2019).

shopping costs (as in Pozzi (2012) and Thomassen et al. (2017)) and preference for variety (as in Hendel (1999) and Dubé (2004)) represent two likely alternatives.

Our model can rationalize shopping costs with positive demand synergies and, as we show in Appendix 8.1, preference for variety with some additional structure on the demand synergy parameters. We try to distinguish empirically the relative contribution of these mechanisms to the estimated demand synergies. Our results show that demand for RTE cereals exhibits substantial Hicksian complementarity and that around 75% of it does not seem to be explained by shopping costs or by preference for variety. We compare our estimation results from the full model to those from a model of demand for single brands (similar to Nevo (2000, 2001)) and show that ignoring Hicksian complementarity may result in misleading demand estimates and counterfactuals. Despite the different econometric approach and data, our results are in line with those by Fosgerau et al. (2019), who also document Hicksian complementarity among different RTE cereal brands in the USA.

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 et al. (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 group and family size) and a panel of shopping trips $r = 1, ..., 756663$ 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 83256 of these, RTE cereals are observed to be purchased. We define a market as a store-week combination $t = 1, ..., 1431$.

Over the sample period, the households are observed to purchase 1173 different UPCs of RTE cereals. For feasibility, 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 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 types on the basis of their ingredients: cereal type R refers to "Regular," F/W to "Fiber/Whole Grain," and S to "Added Sugar." Table 7 in Appendix 8.15 lists these RTE cereal brands and their average market shares across the shopping trips with some RTE cereal purchase.[21] 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.

We make the standard assumption that RTE cereal purchases do not determine store choice and take store choice as exogenous in our econometric model. We consider household i to choose the outside option, which we denote by $j = 0$, whenever no RTE cereal brand is purchased during shopping trip r (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.

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 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.[22] For computational convenience, we focus our analysis on the shopping trips with observed purchases of at most two different RTE cereal brands, thus discarding 3.27% of the shopping trips with some RTE cereal purchase. In 17.69% of the shopping trips with some RTE cereal purchase, households are observed to purchase two different brands of RTE cereals.

Table 1 describes how the average bundle size purchased changes among households with different observable characteristics. The top panel of Table 1 shows that larger families are more likely to purchase larger bundles. This accords to the idea of preference for variety by Hendel (1999) and Dubé (2004): in order to satisfy more heterogeneous preferences (e.g., different genders and ages), larger households are more likely to purchase a wider variety of RTE cereal brands on each shopping trip. The central panel of Table 1 highlights the potential relevance of shopping costs, as suggested by Pozzi (2012) and Thomassen et al. (2017): households observed to shop with a higher frequency (facing lower shopping costs) are less likely to purchase bundles of different RTE cereal brands on any shopping trip.[23] The bottom panel of Table 1 divides the households into three income groups and does not suggest any apparent relationship between the level of income and average purchased bundle size.[24]

[21] Market shares are computed over the shopping trips observed in each store-week combination.

[22] For instance, the purchases of different RTE cereal brands across different shopping trips within the same t are considered as independent purchases of single brands rather than bundles. To keep the dimensionality of the problem manageable, 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.

[23] We compute the "weekly shopping frequency" as the average number of shopping trips per week for each household over the entire four-year period of our sample. The median among the 2897 households is 1.80 shopping trips per week.

[24] We create the three income groups on the basis of 12 income classes originally provided in the IRI data, which are ordered in increasing level of income from 1 to 12. We code as "low income" the classes 1-4, "medium income" the classes 5-8, and we group in "high income" the remaining classes 9-12.

Table 1: Average Number of Different Brands per Shopping Trip

	#Households	Ave. Bundle Size
Family Size		
1	732	1.12
2	1184	1.16
≥ 3	981	1.22
Weekly Shopping Frequency		
$(0, 2]$	1779	1.19
$(2, 3]$	810	1.17
> 3	308	1.14
Income Group		
low	679	1.18
medium	1169	1.16
high	1049	1.19

Notes: The Table shows the distribution of family size, weekly shopping frequency, and income group among the 2897 households in our data. See text for the definitions of these variables. For each value of these variables, we report the average number of different RTE cereal brands observed to be purchased per shopping trip by the corresponding households.

We construct choice sets at the level of the store-week t: any household during any shopping trip in t is assumed to face choice set $\mathbf{C}_t$. This is made of three components: single brands, 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 t. Denote this set of available brands by $\mathbf{J}_t$. Households can also purchase bundles $(j_1, j_2) \in (\mathbf{J}_t \times \mathbf{J}_t) \backslash \{(k_1, k_2) | k_1 = k_2\}$ made of pairs of different RTE cereal brands. Finally, households may decide not to purchase any RTE cereal brand, $j = 0$. By combining these purchase possibilities, the choice set faced during all shopping trips in t is $\mathbf{C}_t = \{0\} \cup \mathbf{J}_t \cup (\mathbf{J}_t \times \mathbf{J}_t) \backslash \{(k_1, k_2) | k_1 = k_2\}$.[25]

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 by

[25] The choice set $\mathbf{C}_t$ also excludes those bundles that are never purchased during any of the shopping trips 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.

purchasing brand $j \in \mathbf{J}_t$ during shopping trip r in market t is:

$$\begin{aligned} U_{irtj} &= u_{itj} + \varepsilon_{irtj} \\ &= \delta_{tj} + \mu_{itj} + \varepsilon_{irtj}, \end{aligned} \tag{12}$$

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

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 $d_i^{\alpha}\alpha + v_i$ is a vector of household i-specific price coefficients made of two components: an observable part that is a function of the household characteristics d_i^{α} (to be detailed in the next section) 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 (12) 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 $\Delta_{tj}(p_{tj}, x_{tj}) = 0$ and $\delta_{tj} = \xi_{tj}$. Even though we use household-level data, we face price endogeneity if, for instance, the producer of RTE cereal brand j sets price p_{tj} taking the average utility δ_{tj} into consideration. Our proposed estimator essentially addresses this endogeneity problem by treating the average utility δ_{tj} for each brand j in each market t as a fixed effect.

The indirect utility of i by purchasing bundle $\mathbf{b}$ during shopping trip r in market t is:

$$\begin{aligned} U_{irt\mathbf{b}} &= \sum_{j\in\mathbf{b}} u_{itj} + \Gamma_{i\mathbf{b}} + \varepsilon_{irt\mathbf{b}} \\ &= \sum_{j\in\mathbf{b}} (\delta_{tj} + \mu_{itj}) + \Gamma_{\mathbf{b}} + \zeta_{i\mathbf{b}} + \varepsilon_{irt\mathbf{b}} \\ &= \sum_{j\in\mathbf{b}} \delta_{tj} + \Gamma_{\mathbf{b}} + \left[\sum_{j\in\mathbf{b}} \mu_{itj} + (d_i^{\gamma}\gamma + \tilde{\zeta}_{i\mathbf{b}})\right] + \varepsilon_{irt\mathbf{b}} \\ &= \delta_{t\mathbf{b}} + \mu_{it\mathbf{b}} + \varepsilon_{irt\mathbf{b}}, \end{aligned} \tag{13}$$

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 of $\zeta_{i\mathbf{b}} = d_i^{\gamma}\gamma + \tilde{\zeta}_{i\mathbf{b}}$, where $d_i^{\gamma}\gamma$ is a function of observed household characteristics d_i^{γ} (to be detailed in the next section) and $\tilde{\zeta}_{i\mathbf{b}}$ is an unobserved random component. Because of pure components pricing, i.e. $p_{t\mathbf{b}} = 0$, and the absence of other bundle-specific observed product characteristics, i.e. $x_{t\mathbf{b}} = 0$, we constrain function $g(\cdot,\cdot|\Sigma_g) = 0$ (see Assumption 2).[26] We attempt to empirically distinguish the relative contribution to $\Gamma_{i\mathbf{b}}$ of two alternative mechanisms. In particular, we specify $d_i^{\gamma}\gamma$ to include measures of family size (to proxy for preference for variety) and of average weekly shopping frequency (to proxy for shopping costs).

Finally, the indirect utility of household i by choosing the outside option during shopping trip r in market t is assumed to be:

$$U_{irt0} = \varepsilon_{irt0}. \tag{14}$$

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 $\beta_i = (v_i, \eta_i, \tilde{\zeta}_i) = (v_i, (\eta_{ij})_{j\in\mathbf{J}}, (\tilde{\zeta}_{i\mathbf{b}})_{\mathbf{b}\in\mathbf{B}})$.[27] 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 purchase probability of $\mathbf{b} \in \mathbf{C}_t$ during shopping trip r in market t is:

$$s_{irt\mathbf{b}}(\delta_{t\mathbf{J}_t}, \alpha, \gamma, \Gamma; \beta_i) = \frac{e^{\delta_{t\mathbf{b}}+\mu_{it\mathbf{b}}(\beta_i)}}{\sum_{\mathbf{b}'\in\mathbf{C}_t} e^{\delta_{t\mathbf{b}'}+\mu_{it\mathbf{b}'}(\beta_i)}}. \tag{15}$$

We assume $\beta_i = (v_i, \eta_i, \tilde{\zeta}_i)$ to be normally distributed and denote its c.d.f. by $\Phi(\cdot;\Sigma_F)$. Let $y_{itr\mathbf{b}} \in \{0,1\}$ be an indicator for whether household i purchased $\mathbf{b}$ 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 purchases $y_i = (y_{itr\mathbf{b}})_{t\in T_i, r\in R_{it}, \mathbf{b}\in\mathbf{C}_t}$:

$$L_i(\delta_{1\mathbf{J}_1}, ...\delta_{T\mathbf{J}_T}, \alpha, \gamma, \Gamma, \Sigma_F; y_i) = \int \prod_{t\in T_i} \prod_{r\in R_{it}} \prod_{\mathbf{b}\in\mathbf{C}_t} (s_{irt\mathbf{b}}(\delta_{t\mathbf{J}_t}, \alpha, \gamma, \Gamma; \beta_i))^{y_{itr\mathbf{b}}}\, d\Phi(\beta_i; \Sigma_F). \tag{16}$$

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

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

We estimate the demand parameters $(\delta_{1\mathbf{J}_1}, ...\delta_{T\mathbf{J}_T}, \alpha, \gamma, \Gamma, \Sigma_F)$ on the basis of MLE (11) derived from likelihood function (17).[28] To get a sense of the practical advantages implied by our novel demand inverse, in the current application the proposed MLE reduces the numerical search (with respect to a standard MLE) from 12351 to 133 parameters, i.e. $(\alpha, \gamma, \Gamma, \Sigma_F)$.

[26] We follow Nevo (2000, 2001) in assuming that RTE cereal producers set prices at the brand-level rather than at the bundle-level (i.e., pure components pricing): households purchasing multiple RTE cereal brands during the same shopping trip are assumed to pay the sum of the prices of the single brands.

[27] $\mathbf{J}$ 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 $t = 1, ..., T$.

[28] Even though (17) is expressed in terms of individual purchases $(y_1, ..., y_I)$ rather than of sampled market shares $(\hat{s}_1, ..., \hat{s}_T)$, it can be easily shown that the corresponding MLE satisfies the conditions of Theorem 6.

Restricted Models. In what follows, we refer to the model specified in (15)-(17) as to the *full model* of demand for bundles. To better evaluate the practical relevance of allowing for complementarity, we also estimate two restricted versions of the full model.

In the first restricted model, we constrain $\Gamma_{i\mathbf{b}} = 0$ in estimation for all i's and $\mathbf{b}$'s. A comparison of the full model with this first restricted model highlights the importance of controlling for the demand synergies $\Gamma_{i\mathbf{b}}$ while keeping everything else unchanged. Despite the absence of demand synergies, this restricted model can still give rise Hicksian complementarity.[29]

In the second restricted model, we rule out Hicksian complementarity by constraining $\Gamma_{i\mathbf{b}} = -\infty$ in estimation for all i's and $\mathbf{b}$'s. In other words, choice sets are restricted not to include any bundle. This amounts to specifying a standard model of demand for single RTE cereal brands (along the lines of Nevo (2000, 2001)) with choice set $\mathbf{C}_t = \{0\} \cup \mathbf{J}_t$ in each t. This second restricted model is estimated on the basis of the same purchase observations as the other two models. However, the observations are used differently: the second restricted model does not differentiate between simultaneous (during the same shopping trip) and sequential (during different shopping trips) purchases of different brands.

6.3 Estimates of Demand for RTE Cereal Bundles

In this section, we present our estimation results for the full model. We postpone a comparison of the estimation results from the three different models to the next two sections, where we discuss price elasticities and counterfactual simulations.

We capture observed heterogeneity in price sensitivity $-\exp(d_i^\alpha \alpha + v_i)$ by d_i^α, a vector of nine mutually exclusive dummies indicating household i's income group (low, medium, and high) and family size (one, two, and larger than two). Unobserved heterogeneity in price sensitivity is instead captured by the random coefficient v_i, which we assume to be i.i.d. normal with standard deviation σ_v. For each of the three estimated models, the top panel of Table 2 reports estimates of the average price sensitivity $\mathbb{E}[-\exp(d_i^\alpha \alpha + v_i)|d_i^\alpha]$ for each value of d_i^α. These results do not seem to suggest any systematic heterogeneity in price sensitivity among households with different incomes and family sizes.

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. Remember that cereal type R refers to "Regular," F/W to "Fiber/Whole Grain," and S to "Added Sugar." Note that any two brands with similar ingredients will share the same $\eta_{i,\text{type}}$. We assume $\tilde{\eta}_{ij}$ to be distributed normal with

[29] In this first restricted model, the cross-price elasticities can still be negative because the choice set $\mathbf{C}_t = \{0\} \cup \mathbf{J}_t \cup (\mathbf{J}_t \times \mathbf{J}_t) \backslash \{(k_1, k_2) | k_1 = k_2\}$ is not *complete*. $\mathbf{C}_t$ would be complete if it included also the bundles made of two units of the same brand. Gentzkow (2007)'s Proposition 1 at page 719, which states that a positive demand synergy is necessary and sufficient for Hicksian complementarity, only applies to models with complete choice sets.

standard deviation $\sigma_{\tilde{\eta}}$. Following Gentzkow (2007), we also assume $(\eta_{i,R}, \eta_{i,F/W}, \eta_{i,S})$ to be distributed jointly normal with standard deviations and pairwise correlations denoted by, respectively, $\sigma_{\eta_{\text{type}}}$ and $\text{corr}_{\text{type,type}'}$, type, type$' \in \{R, F/W, S\}$. We allow single-person households and multi-person households to have different joint normal distributions of $(\eta_{i,R}, \eta_{i,F/W}, \eta_{i,S})$. We assume $\tilde{\eta}_{ij}$ and $\eta_{i,\text{type}}$ to be mutually independent.

The estimates of the distribution of η_{ij} and of the other random coefficients are reported in the central panel of Table 2. Overall, the estimates are highly significant and underline the importance of controlling for unobserved heterogeneity, not only in terms of price sensitivity, but also of brand-specific random intercepts and of demand synergies (we return to these in more detail below). 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 cereal brands. Households of different family sizes do not seem to have systematically different distributions of $(\eta_{i,R}, \eta_{i,F/W}, \eta_{i,S})$.

We specify the demand synergy of household i for bundle $\mathbf{b}$ as:

$$\begin{aligned}\Gamma_{i\mathbf{b}} =& \Gamma_{\mathbf{b}} + d_i^{\gamma}\gamma + \tilde{\zeta}_{i\mathbf{b}} \\ =& \Gamma_{\mathbf{b}} + \gamma_2 \mathbf{1}\{\text{family size}_i = 2\} + \gamma_{\geq 3}\mathbf{1}\{\text{family size}_i \geq 3\} \\ & + \gamma_s \mathbf{1}\{\text{normal shopping frequency}_i\} + \tilde{\zeta}_{i\mathbf{b}},\end{aligned} \tag{18}$$

where $\mathbf{1}\{\cdot\}$ is the indicator function and "normal shopping frequency$_i$" denotes whether the average weekly shopping frequency of household i lies below the 95^{th} percentile.[30] Parameter γ_k captures systematic differences between the average demand synergies of households of family size k and single-person households. We include family size in the specification of $\Gamma_{i\mathbf{b}}$ as a proxy for preference for variety. γ_s instead measures differences in the average demand synergies between households observed to shop at a normal frequency, i.e. in the bottom 95% of the distribution, and households who shop very often, i.e. in the top 5% of the distribution. We control for *normal* shopping frequency in (18) as a proxy for *larger* shopping costs. These are meant to rationalize the purchase patterns documented in Table 1: larger families may have to satisfy more heterogeneous preferences within the household, while more frequent shoppers may be less likely to purchase multiple brands on any shopping trip. $\tilde{\zeta}_{i\mathbf{b}}$ represents a i-specific unobserved component of demand synergy for bundle $\mathbf{b}$, which we assume to be i.i.d. normal with standard deviation $\sigma_{\tilde{\zeta}}$. $\tilde{\zeta}_{i\mathbf{b}}$ allows for the possibility that the brands in bundle $\mathbf{b}$ have positive demand synergies for some households and negative for others. Estimates of the $\Gamma_{\mathbf{b}}$'s are reported in Table 3, while estimates of the remaining demand synergy parameters are reported in the bottom panel of Table 2.

[30] In our sample of households, the 95^{th} percentile of the average weekly shopping frequency (i.e., the average number of shopping trips in a week) is 3.67.

Table 2: Demand Estimates for Full and Restricted Models

	Full Model	Restricted Model 1 $\Gamma_{i\mathbf{b}} = 0$	Restricted Model 2 $\Gamma_{i\mathbf{b}} = -\infty$
Average Price Sensitivities			
low income, family size = 1	−0.44 (0.164)	−0.49 (0.196)	−1.27 (0.180)
family size = 2	−0.47 (0.175)	−0.51 (0.202)	−1.30 (0.185)
family size ≥ 3	−0.39 (0.147)	−0.38 (0.151)	−1.17 (0.167)
medium income, family size = 1	−0.47 (0.174)	−0.52 (0.210)	−1.29 (0.183)
family size = 2	−0.42 (0.157)	−0.46 (0.183)	−1.27 (0.180)
family size ≥ 3	−0.49 (0.183)	−0.49 (0.195)	−1.31 (0.185)
high income, family size = 1	−0.39 (0.146)	−0.43 (0.173)	−1.20 (0.170)
family size = 2	−0.42 (0.157)	−0.45 (0.180)	−1.26 (0.179)
family size ≥ 3	−0.40 (0.150)	−0.41 (0.165)	−1.25 (0.177)
Random Coefficients			
price, σ_v	0.36 (0.004)	0.39 (0.004)	0.14 (0.002)
brand intercepts, $\sigma_{\tilde{\eta}}$	0.87 (0.005)	0.87 (0.005)	0.81 (0.005)
demand synergies, $\sigma_{\tilde{\zeta}}$	0.06 (0.010)		
Single-Person Households			
σ_{η_R}	0.50 (0.037)	0.47 (0.046)	0.20 (0.037)
$\sigma_{\eta_{F/W}}$	0.54 (0.014)	0.52 (0.024)	0.65 (0.017)
σ_{η_S}	0.96 (0.030)	0.90 (0.034)	0.97 (0.030)
$\text{corr}_{R,F/W}$	−0.86 (0.014)	−0.84 (0.016)	−0.89 (0.012)
$\text{corr}_{R,S}$	−0.52 (0.042)	−0.60 (0.044)	−0.61 (0.036)
$\text{corr}_{F/W,S}$	0.29 (0.042)	0.41 (0.056)	0.47 (0.044)
Multi-Person Households			
σ_{η_R}	0.10 (0.016)	0.37 (0.016)	0.45 (0.015)
$\sigma_{\eta_{F/W}}$	0.54 (0.008)	0.70 (0.008)	0.86 (0.009)
σ_{η_S}	0.96 (0.017)	1.10 (0.017)	1.24 (0.015)
$\text{corr}_{R,F/W}$	−0.93 (0.006)	−0.91 (0.006)	−0.95 (0.004)
$\text{corr}_{R,S}$	−0.79 (0.016)	−0.73 (0.015)	−0.78 (0.010)
$\text{corr}_{F/W,S}$	0.73 (0.019)	0.81 (0.011)	0.85 (0.008)
Demand Synergies, γ			
family size= 2, γ_2	0.14 (0.017)		
family size$\geq$ 3, $\gamma_{\geq 3}$	0.36 (0.015)		
normal shop. freq., γ_s	0.001 (0.012)		

Notes: Each column of the Table reports estimates from one of three model specifications: the full model, restricted model 1 (which constrains $\Gamma_{i\mathbf{b}} = 0$ in estimation), and restricted model 2 (which constrains $\Gamma_{i\mathbf{b}} = -\infty$ in estimation, i.e. standard demand model for single brands). The top panel reports the estimated average price sensitivity $\mathbb{E}[-\exp(d_i^\alpha \alpha + v_i)|d_i^\alpha]$ for each value of d_i^α and the corresponding standard deviation (in brackets), computed as $\sqrt{\text{Var}[\exp(d_i^\alpha \alpha + v_i)|d_i^\alpha]}$. The central panel reports estimates of the parameters characterizing the distribution of the random coefficients, while the bottom panel those of the demand synergy parameters associated to different family sizes and weekly shopping frequencies. For the estimates in the central and bottom panel, standard errors are reported in brackets. Cereal type R refers to "Regular," F/W to "Fiber/Whole Grain," and S to "Added Sugar."

The estimates from Table 3 suggest that several pairs of RTE cereal brands have positive $\Gamma_{\mathbf{b}}$'s.[31] Importantly for competition policy, as we will explore in the next section, there appear to be positive $\Gamma_{\mathbf{b}}$'s not only among brands within the same producer, but also among brands sold by different producers. For example, the first column of Table 3 shows that single-person households exhibit positive demand synergies between General Mills and most of the 15 remaining brands. Moreover, in line with the evidence from Table 1, the estimated demand synergy shifters γ_2 and $\gamma_{\geq 3}$ from the bottom panel of Table 2 are positive and increasing in family size, i.e. $\gamma_{\geq 3} \geq \gamma_2$. We interpret this as evidence of preference for variety: larger families exhibit more positive demand synergies among different RTE cereal brands than smaller families. Intuitively, larger families may be more likely to purchase different brands in order to satisfy more heterogeneous RTE cereal tastes within the household (e.g., adults and children of different ages). Differently, γ_s is positive but not significantly different from zero, highlighting that—after controlling for everything else—households with different shopping frequencies are similarly likely to purchase bundles of different brands on any shopping trip. The standard deviation $\sigma_{\tilde{\zeta}}$ of the random coefficient $\tilde{\zeta}_{i\mathbf{b}}$ is estimated to be small but significant, suggesting the presence of household-specific heterogeneity in demand synergies beyond differences in family size and weekly shopping frequency.

Evidence in Support of Assumption 2. As discussed in Remark 1, Assumption 2 can be verified in practice. In Appendix 8.7, we present a specification test for Assumption 2 that builds on partial identification methods. A rejection of the test is evidence against Assumption 2. In the context of our empirical illustration, the test statistic evaluated at the estimates from Tables 2 and 3 is 9910, which is smaller than the critical value for rejection at the 10% level, 21081 (a chi-square with 20819 degrees of freedom). This strongly suggests that $\Theta_I(\mathbf{T})$ in (31) is not empty, providing reassuring evidence in support of Assumption 2.

6.4 Hicksian Complementarity and Demand Synergies

Table 4 reports the average (across markets) estimated own- and cross-price elasticities of demand from the full model. Each entry reports the percent change in the brand-level 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 $\mathbf{b} \in \mathbf{C}_{t1}$ in market t, the estimated brand-level market share function of brand $j \in \mathbf{J}_t$ is defined as $\hat{s}_{tj.} = \sum_{\mathbf{b} \in \mathbf{C}_{t1}: j \in \mathbf{b}} \hat{s}_{t\mathbf{b}}$. Table 4 provides pervasive evidence of Hicksian complementarity. For example, the first column shows that households exhibit statistically significant complementarity between General Mills and several of the 15 remaining brands. According to intuition, Hicksian complementarity seems to be more pronounced among those brands with larger positive $\Gamma_{\mathbf{b}}$ (see Table 3).

[31]Note that these estimates come from the full model, neither of the restricted models allows for demand synergies. See the Table notes for an interpretation of the missing values.

In our specification, Hicksian complementarity among different RTE cereal brands can be explained by alternative mechanisms: correlation in the unobserved preferences for single brands (η_{ij}), preference for variety (γ_2 and $\gamma_{\geq 3}$), shopping costs (γ_s), bundle-specific fixed effects ($\Gamma_{\mathbf{b}}$'s)—which, among other things, may account for synergies in consumption—, and residual unobserved heterogeneity ($\tilde{\zeta}_{t\mathbf{b}}$). To shed light on the relative contributions of these mechanisms, we sequentially "switch them off" from the estimated full model and re-compute the cross-price elasticities. Table 5 summarizes the results.

Different from Gentzkow (2007), the unobserved preferences for single brands (η_{ij}) contribute to the substitutability among RTE cereal brands (−2.00%), possibly because of the negative correlation between $\eta_{i,R}$ and both $\eta_{i,F/W}$ and $\eta_{i,S}$ (see Table 2). The average of the cross-price elasticities instead increases (becoming less negative) as we progressively switch off the various components of $\Gamma_{i\mathbf{b}}$: residual unobserved heterogeneity (+0.09%), shopping costs (+0.05%), and especially preference for variety (+15.22%).

However, the most dramatic changes occur when we further set the bundle-specific fixed effects $\Gamma_{\mathbf{b}}$'s either to zero (+54.94%) or to $-\infty$ (+31.69%). While this is expected in the case of $\Gamma_{i\mathbf{b}} = -\infty$ (standard demand model for single brands), the average of the cross-price elasticities already changes from negative (complementarity) to positive (substitutability) when setting each $\Gamma_{i\mathbf{b}}$ to zero. Collectively, these results suggest that most of the estimated complementarity is explained by preference for variety and by the bundle-specific fixed effects.

Standard models of demand for single brands rule out Hicksian complementarity among different RTE cereal brands and restrict the cross-price elasticities to be positive. Ignoring the presence of complementarity among different brands may lead to incorrect demand estimates and misleading price elasticities. In order to quantify the extent of this problem, we compare the price elasticities computed on the basis of the estimates from the full model (Table 4) to those computed on the basis of the estimates from the restricted models (Supplement Tables 8 and 9).[32] Several of the estimated cross-price elasticities have opposite signs, mistakenly suggesting substitutability rather than complementarity among different pairs of RTE cereal brands. To further explore the economic consequences of accounting for complementarity in demand estimation, in the next section we compare some counterfactual simulations implied by estimates from the full model to those implied by the estimates from the restricted models.

[32]Demand estimates from the full model can be found in the first column of Table 2 and in Table 3, while those from the restricted models can be found in the second and third columns of Table 2 (the restricted models do not include demand synergies).

Table 3: 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	0.93 (0.036)	·														
	F/W	0.89 (0.027)	1.83 (0.034)	·													
	S	0.98 (0.051)	2.12 (0.059)	2.23 (0.041)	·												
Quaker	R	0.80 (0.851)	1.82 (0.815)	- 0.97 (0.992)	1.50 (0.935)	·											
	F/W	0.90 (0.050)	1.17 (0.093)	0.95 (0.067)	0.84 (0.295)	0.35 (0.992)	·										
	S	0.96 (0.081)	1.05 (0.522)	1.10 (0.110)	1.41 (0.344)	2.72 (0.918)	2.52 (0.090)	·									
Post	R	1.66 (0.985)	2.22 (0.983)	2.04 (0.986)	·	·	3.62 (0.988)	·	·								
	F/W	0.97 (0.036)	0.95 (0.072)	0.89 (0.050)	0.81 (0.107)	1.76 (0.925)	1.30 (0.079)	1.05 (0.273)	4.01 (0.964)	·							
	S	0.96 (0.145)	0.56 (0.908)	1.19 (0.298)	0.78 (0.855)	·	0.86 (0.931)	1.49 (0.912)	·	1.97 (0.208)	·						
Private	R	0.96 (0.059)	0.53 (0.252)	0.62 (0.097)	0.08 (0.674)	1.41 (0.971)	0.99 (0.257)	0.49 (0.833)	·	0.90 (0.115)	0.20 (0.947)	·					
	F/W	0.49 (0.067)	0.56 (0.318)	0.40 (0.098)	0.27 (0.731)	·	1.79 (0.104)	0.12 (0.949)	4.45 (1.255)	0.67 (0.245)	0.46 (0.942)	2.31 (0.082)	·				
	S	0.57 (0.084)	0.79 (0.370)	0.73 (0.115)	- 0.01 (0.446)	1.51 (0.992)	0.84 (0.311)	0.88 (0.527)	·	0.72 (0.395)	0.61 (0.794)	2.83 (0.081)	2.63 (0.091)	·			
Small P.	R	0.21 (0.143)	- 0.01 (0.723)	0.54 (0.357)	0.35 (0.877)	1.66 (0.990)	0.40 (0.850)	0.68 (0.912)	·	0.19 (0.799)	·	0.71 (0.725)	0.26 (0.066)	0.01 (0.893)	·		
	F/W	0.96 (0.033)	0.50 (0.078)	0.39 (0.059)	0.21 (0.171)	·	0.32 (0.301)	0.34 (0.659)	·	0.52 (0.074)	0.16 (0.895)	0.54 (0.227)	0.97 (0.096)	0.99 (0.389)	2.72 (0.066)	·	
	S	0.23 (0.544)	- 0.36 (0.908)	0.54 (0.532)	- 0.20 (0.890)	·	1.39 (0.720)	0.17 (0.930)	·	0.64 (0.707)	0.95 (0.974)	1.56 (0.819)	1.38 (0.933)	1.50 (0.567)	1.80 (0.893)	2.90 (0.096)	·

Notes: The Table reports the estimated average demand synergies for all bundles of RTE cereal brands for single-person households (standard errors in brackets). 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.

Table 4: Average Estimated Own- and Cross-Price Elasticities (Full Model)

		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	-1.226	-0.016	-0.022	-0.007	-0.002	-0.006	-0.002	-0.001	-0.022	-0.001	-0.006	-0.001	-0.001	0.001	-0.024	0.001
		(0.154)	(0.002)	(0.003)	(0.001)	(0.007)	(0.001)	(0.001)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.004)	(0.002)
Kellogg's	R	-0.047	-1.387	-0.073	-0.033	-0.010	-0.006	-0.001	-0.002	-0.012	0.000	-0.001	0.000	0.000	0.003	0.001	0.001
		(0.007)	(0.184)	(0.011)	(0.004)	(0.017)	(0.002)	(0.006)	(0.002)	(0.003)	(0.002)	(0.010)	(0.007)	(0.011)	(0.012)	(0.002)	(0.003)
	F/W	-0.053	-0.060	-1.152	-0.048	0.003	-0.004	-0.002	-0.001	-0.012	-0.001	0.000	0.001	0.000	0.000	0.005	0.001
		(0.007)	(0.009)	(0.147)	(0.006)	(0.001)	(0.001)	(0.001)	(0.004)	(0.003)	(0.004)	(0.001)	(0.001)	(0.002)	(0.007)	(0.002)	(0.003)
	S	-0.047	-0.074	-0.133	-1.134	-0.003	0.000	-0.007	0.001	-0.004	0.001	0.004	0.003	0.007	0.002	0.013	0.004
		(0.009)	(0.010)	(0.018)	(0.146)	(0.010)	(0.011)	(0.012)	(0.000)	(0.009)	(0.003)	(0.004)	(0.010)	(0.006)	(0.007)	(0.013)	(0.003)
Quaker	R	−0.036	-0.067	0.023	-0.012	-1.538	0.002	-0.022	.	-0.054	0.001	-0.013	0.005	-0.007	-0.014	0.027	0.003
		(0.135)	(0.114)	(0.010)	(0.036)	(0.230)	(0.024)	(0.033)		(0.081)	(0.000)	(0.019)	(0.001)	(0.022)	(0.032)	(0.004)	(0.001)
	F/W	-0.050	-0.017	-0.014	0.000	0.001	-1.126	-0.024	-0.010	-0.030	0.000	-0.004	-0.019	0.000	0.002	0.009	-0.002
		(0.010)	(0.007)	(0.005)	(0.014)	(0.011)	(0.144)	(0.004)	(0.009)	(0.006)	(0.004)	(0.014)	(0.006)	(0.007)	(0.018)	(0.033)	(0.019)
	S	-0.048	-0.009	-0.017	-0.021	-0.029	-0.059	-1.196	0.001	-0.016	-0.002	0.003	0.004	-0.001	0.000	0.011	0.003
		(0.014)	(0.038)	(0.012)	(0.036)	(0.042)	(0.010)	(0.155)	(0.000)	(0.042)	(0.009)	(0.015)	(0.009)	(0.028)	(0.009)	(0.007)	(0.004)
Post	R	-0.054	-0.024	-0.038	0.009	.	-0.149	0.001	-0.861	-0.223	.	0.003	-0.058	.	0.003	0.019	0.006
		(0.102)	(0.028)	(0.098)	(0.002)		(0.131)	(0.000)	(0.110)	(0.190)		(0.001)	(0.113)		(0.000)	(0.003)	(0.001)
	F/W	-0.072	-0.013	-0.017	-0.002	-0.009	-0.011	-0.002	-0.017	-1.160	-0.005	-0.004	-0.002	-0.001	0.002	-0.002	0.000
		(0.012)	(0.003)	(0.004)	(0.004)	(0.014)	(0.002)	(0.006)	(0.015)	(0.146)	(0.005)	(0.003)	(0.011)	(0.009)	(0.007)	(0.002)	(0.004)
	S	-0.044	0.007	-0.017	0.005	0.003	0.000	-0.006	.	-0.118	-1.111	0.005	0.001	0.002	0.007	0.019	0.000
		(0.047)	(0.035)	(0.070)	(0.031)	(0.000)	(0.023)	(0.020)		(0.106)	(0.137)	(0.018)	(0.006)	(0.013)	(0.001)	(0.044)	(0.007)
Private	R	-0.059	-0.003	-0.002	0.006	-0.006	-0.004	0.001	0.001	-0.012	0.001	-0.968	-0.029	-0.043	-0.001	-0.001	-0.003
		(0.010)	(0.034)	(0.005)	(0.007)	(0.008)	(0.016)	(0.007)	(0.000)	(0.010)	(0.003)	(0.120)	(0.005)	(0.006)	(0.006)	(0.021)	(0.008)
	F/W	-0.009	0.001	0.006	0.006	0.003	-0.027	0.003	-0.028	-0.006	0.001	-0.037	-0.900	-0.042	0.002	-0.016	-0.003
		(0.009)	(0.029)	(0.006)	(0.020)	(0.001)	(0.009)	(0.006)	(0.054)	(0.042)	(0.002)	(0.007)	(0.109)	(0.006)	(0.012)	(0.008)	(0.012)
	S	-0.012	-0.003	-0.002	0.015	-0.005	-0.001	-0.001	.	-0.004	0.001	-0.067	-0.046	-0.893	0.003	-0.014	-0.005
		(0.009)	(0.054)	(0.011)	(0.013)	(0.015)	(0.012)	(0.018)		(0.042)	(0.004)	(0.010)	(0.007)	(0.113)	(0.005)	(0.043)	(0.009)
Small P.	R	0.011	0.011	0.000	0.003	-0.007	0.002	0.000	0.001	0.006	0.002	-0.002	0.001	0.002	-1.572	-0.209	-0.004
		(0.037)	(0.045)	(0.032)	(0.012)	(0.015)	(0.024)	(0.005)	(0.000)	(0.026)	(0.000)	(0.007)	(0.011)	(0.004)	(0.206)	(0.034)	(0.016)
	F/W	-0.072	0.001	0.006	0.006	0.004	0.003	0.002	0.001	-0.002	0.001	0.000	-0.004	-0.003	-0.052	-1.333	-0.022
		(0.011)	(0.002)	(0.003)	(0.006)	(0.001)	(0.011)	(0.004)	(0.000)	(0.002)	(0.002)	(0.006)	(0.002)	(0.009)	(0.008)	(0.173)	(0.005)
	S	0.030	0.012	0.012	0.020	0.003	-0.008	0.005	0.001	0.002	0.000	-0.010	-0.006	-0.012	-0.011	-0.249	-0.917
		(0.079)	(0.033)	(0.049)	(0.014)	(0.001)	(0.073)	(0.007)	(0.000)	(0.035)	(0.004)	(0.031)	(0.028)	(0.023)	(0.040)	(0.055)	(0.113)

Notes: The Table reports the average estimated own- (diagonal) and cross-price (off-diagonal) elasticities from the full model, where averages are computed across markets. Each entry reports the percent change in the brand-level 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 brand-level 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." The missing values "·" refer to situations where the corresponding column and row brands are never simultaneously sold in the same market. The standard errors of the estimated elasticities are in brackets obtained from a parametric bootstrap as in Nevo (2000, 2001) with 50 draws.

Table 5: Cross-Price Elasticities and Demand Synergies

	Average	Relative Change
Estimated Full Model	-0.0126	–
Estimated Full Model, then switch off η_{ij}	-0.0130	-2.00%
Estimated Full Model, then switch off η_{ij} and $\tilde{\zeta}_{i\mathbf{b}}$	-0.0130	0.09%
Estimated Full Model, then switch off η_{ij}, $\tilde{\zeta}_{i\mathbf{b}}$, and γ_s	-0.0130	0.05%
Estimated Full Model, then switch off η_{ij}, $\tilde{\zeta}_{i\mathbf{b}}$, γ_s, γ_2, and $\gamma_{\geq 3}$	-0.0100	15.22%
Estimated Full Model, switch off η_{ij} and set $\Gamma_{i\mathbf{b}} = 0$	0.0007	54.94%
Estimated Full Model, switch off η_{ij} and set $\Gamma_{i\mathbf{b}} = -\infty$	0.0068	31.69%

Note: The Table reports the average (across markets) of the cross-price elasticities of all pairs of different RTE cereal brands. All the cross-price elasticities are obtained from the full model, where the estimated parameters in the specification of $\Gamma_{i\mathbf{b}}$ from (18) are "switched off" sequentially. For example, the third row is obtained from the estimated full model by setting the standard deviations of η_{ij} and $\tilde{\zeta}_{i\mathbf{b}}$ to zero, while the fourth is obtained by further setting the estimated γ_s to zero. The column "Relative Change" reports the percent change in the average of the cross-price elasticities from any two consecutive rows: for instance, the relative change from the fourth row to the fifth is 15.22%.

6.5 Counterfactuals and Comparisons with Standard Model

Here we evaluate the economic relevance of allowing for Hicksian complementarity by comparing some counterfactuals from the full model, with $\Gamma_{i\mathbf{b}}$ as in (18), to those from two restricted models: restricted model 1, which constrains $\Gamma_{i\mathbf{b}} = 0$ in estimation, and restricted model 2—the standard demand model for single brands—, which constrains $\Gamma_{i\mathbf{b}} = -\infty$ in estimation thereby forcing substitutability. For the counterfactuals, we take the observed scenario of pure components pricing and oligopolistic competition among RTE cereal producers as a reference (see Nevo (2000, 2001) for the institutional details), and simulate the changes in prices, profits, and consumer surplus implied by different market structures.[33]

The results of these counterfactuals are reported in Table 6. The Table reports relative changes in prices (top panel), profits (central panel), and consumer surplus (bottom panel) associated with each of three counterfactual market structures (columns) as simulated by each of the three estimated models (rows). We consider four alternative market structures: "competition," where we suppose that each single brand is owned and sold by a different (fictional) producer (for a total of 16 producers); "oligopoly," which corresponds to the observed oligopolistic competition among six producers; "duopoly," where we suppose that five of the producers (General Mills, Kellogg'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,"

[33]Given our estimates of demand and marginal costs, we simulate each profile of counterfactual prices—independently for each market—using the necessary first order conditions for a Nash equilibrium of the corresponding pure components pricing game. For example, in a monopoly, the same agent chooses a specific price for each single brand so to maximize industry profits.

where we suppose that the six producers perfectly collude as a monopolist.

The simulation results from the full model confirm the classic insight by Cournot (1838): mergers between producers selling complementary brands can be socially desirable. In pure components pricing, the prices of all single brands—and consequently of all bundles—decrease as the level of competition weakens: while industry-level profit remains basically unchanged, consumer surplus increases with market concentration. As market structure becomes more concentrated, producers internalize more of the externalities due to complementarity and consequently choose lower prices, as can be seen from the relative increase in consumer surplus from -5.27% for competition to $+7.62\%$ for monopoly.

Strikingly, the restricted models produce opposite predictions: as market structure becomes more concentrated, prices increase and consumer surplus decreases. As shown in Supplement Tables 8 and 9, both restricted models predict positive cross-price elasticities and therefore substitutability among any pair of RTE cereal brands. Not surprisingly then, any merger between producers selling substitutable brands will lead to higher prices and ultimately hurt consumers. These results underline the economic importance of allowing for both substitutability and complementarity in demand estimation: while estimates from the full model provide supportive evidence for the classic Cournot (1838)'s insight that mergers can be welfare enhancing, those from a standard model that does not allow for Hicksian complementarity can only predict that mergers will be detrimental for consumer surplus.

Table 6: Counterfactual Simulations

	Competition	Oligopoly	Duopoly	Monopoly
Price change				
Full Model	+8.08% (1.35%)	0%	−5.03% (1.16%)	−5.34% (1.65%)
Restricted Model 1, $\Gamma_{i\mathbf{b}} = 0$	−0.54% (0.03%)	0%	+3.69% (0.21%)	+5.26% (0.30%)
Restricted Model 2, $\Gamma_{i\mathbf{b}} = -\infty$	−0.56% (0.10%)	0%	+4.17% (0.72%)	+6.06% (1.07%)
Profit change				
Full Model	−0.47% (0.04%)	0%	+0.27% (0.03%)	+0.30% (0.05%)
Restricted Model 1, $\Gamma_{i\mathbf{b}} = 0$	−0.03% (0.002%)	0%	+0.10% (0.003%)	+0.12% (0.004%)
Restricted Model 2, $\Gamma_{i\mathbf{b}} = -\infty$	−0.36% (0.01%)	0%	+1.64% (0.03%)	+2.00% (0.03%)
Consumer Surplus change				
Full Model	−5.27% (0.22%)	0%	+6.29% (0.37%)	+7.62% (0.62%)
Restricted Model 1, $\Gamma_{i\mathbf{b}} = 0$	+0.49% (0.01%)	0%	−3.23% (0.04%)	−4.54% (0.06%)
Restricted Model 2, $\Gamma_{i\mathbf{b}} = -\infty$	+1.53% (0.01%)	0%	−11.94% (0.07%)	−16.41% (0.09%)

Notes: The Table reports average counterfactual changes in prices (top panel), profits (central panel), and consumer surplus (bottom panel) of pure components pricing under alternative simulated market structures with respect to the observed oligopoly. Each column refers to a specific market structure: the second column refers to the observed oligopoly in the data while the others refer to simulated counterfactuals (see text for details). Each row refers to one of three model specifications: the full model, restricted model 1 (which constrains $\Gamma_{i\mathbf{b}} = 0$ in estimation), and restricted model 2 (which constrains $\Gamma_{i\mathbf{b}} = -\infty$ in estimation, i.e. standard demand model for single brands). The standard errors associated to the estimated relative changes are in brackets and obtained from a parametric bootstrap as in Nevo (2000, 2001) with 50 draws.

7 Conclusions

We present a novel identification and estimation strategy of a mixed logit model of demand for bundles with endogenous prices given observations on bundle-level market shares. We propose a novel demand inverse in the presence of complementarity that allows to concentrate out of the likelihood function the (potentially numerous) market-product specific average utilities and to substantially alleviate the challenge of dimensionality inherent in estimation. Finally, we estimate demand and supply in the US ready-to-eat cereal industry, where our estimator reduces the numerical search from approximately 12000 to 130 parameters. Our results suggest that ignoring Hicksian complementarity among products often purchased in bundles may result in misleading demand estimates and counterfactuals.

Our identification and estimation arguments are developed for mixed logit models with parametric distributions of random coefficients. In light of the well known challenge of dimensionality that affects the estimation of demand for bundles (Berry et al. (2014)), our priority is to propose estimation methods that can be practically useful in applications that involve more than a few products. While our parametric assumptions clearly help in alleviating the challenge of dimensionality in estimation, they require the econometrician to take a stand on the functional form of the distribution of random coefficients. An avenue for future research is the study of the semi-parametric identification of mixed logit models of demand for bundles, where the distribution of random coefficients is allowed to be non-parametric and more robust against misspecification.

The implementation of our methods requires the observation of bundle-level market shares rather than of the more readily available aggregate market shares of single products. Even though direct measures of bundle-level market shares are widely available only for a few industries, such as media and telecommunication (Crawford and Yurukoglu (2012) and Crawford et al. (2018)), it is usually possible to construct indirect measures of bundle-level market shares from samples of household-level purchases (Gentzkow (2007), Kwak et al. (2015), Grzybowski and Verboven (2016), Ruiz et al. (2017), and Ershov et al. (2018)). In some important industries, however, only measures of aggregate market shares of single products are widely available (e.g., the car industry, see Berry et al. (1995, 2004a)) even though households are known to purchase bundles of products (Manski and Sherman (1980)). When only aggregate market shares of single products are available, our proposed methods do not apply. An important direction for future research is thus the identification and estimation of models of demand for bundles on the basis of aggregate market shares of single products (see Sher and Kim (2014), Allen and Rehbeck (2019), and Wang (2019)).

References

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

Allen, R. and Rehbeck, J. (2019). Identification with additively separable heterogeneity. *Econometrica*, 87(3):1021–1054.

Amemiya, T. (1985). *Advanced econometrics*. Harvard university press.

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

Armstrong, M. (2016a). Nonlinear pricing. *Annual Review of Economics*, 8:583–614.

Armstrong, M. and Vickers, J. (2010). Competitive non-linear pricing and bundling. *The Review of Economic Studies*, 77(1):30–60.

Armstrong, T. B. (2016b). Large market asymptotics for differentiated product demand estimators with economic models of supply. *Econometrica*, 84(5):1961–1980.

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

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

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

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890.

Berry, S., Levinsohn, J., and Pakes, A. (2004a). Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of political Economy*, 112(1):68–105.

Berry, S., Linton, O. B., and Pakes, A. (2004b). Limit theorems for estimating the parameters of differentiated product demand systems. *The Review of Economic Studies*, 71(3):613–654.

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

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

Boothby, W. M. (1986). *An introduction to differentiable manifolds and Riemannian geometry*, volume 120. Academic press.

Bowden, R. (1973). The theory of parametric identification. *Econometrica: Journal of the Econometric Society*, pages 1069–1074.

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

Chan, T. Y. (2006). Estimating a continuous hedonic-choice model with an application to demand for soft drinks. *The Rand journal of economics*, 37(2):466–482.

Chernozhukov, V., Hong, H., and Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, 75(5):1243–1284.

Chesher, A. and Rosen, A. M. (2017). Generalized instrumental variable models. *Econometrica*, 85(3):959–989.

Chu, C. S., Leslie, P., and Sorensen, A. (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*, volume 130. Siam.

Compiani, G. (2019). Market counterfactuals and the specification of multi-product demand: A nonparametric approach. *Working Paper*.

Conlon, C. and Gortmaker, J. (2019). Best practices for differentiated products demand estimation with pyblp. Technical report, Working paper. url: https://chrisconlon. github. io/site/pyblp. pdf.

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

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

Crawford, G. S. and Yurukoglu, A. (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*. Number 17. Yale University Press.

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

Dubé, J.-P. H. (2019). Microeconometric models of consumer demand. *Handbook on the Economics of Marketing.*

Ershov, D., Laliberté, W. P., and Orr, S. (2018). Mergers in a model with complementarity. *Working Paper*, pages 1–50.

Fan, Y. (2013). Ownership consolidation and product characteristics: A study of the us daily newspaper market. *American Economic Review*, 103(5):1598–1628.

Fosgerau, M., Monardo, J., and de Palma, A. (2019). The inverse product differentiation logit model. *Working Paper.*

Fox, J. T., Kim, K., Ryan, S. P., and Bajari, P. (2012). The random coefficients logit model is identified. *Journal of Econometrics*, 166(2):204–212.

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

Freyberger, J. (2015). Asymptotic theory for differentiated products demand models with many markets. *Journal of Econometrics*, 185(1):162–181.

Gale, D. and Nikaido, H. (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.

Gentzkow, M., Shapiro, J. M., and Sinkinson, M. (2014). Competition and ideological diversity: Historical evidence from us newspapers. *American Economic Review*, 104(10):3073–3114.

Grzybowski, L. and Verboven, F. (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.

Ho, K., Ho, J., and Mortimer, J. H. (2012). The use of full-line forcing contracts in the video rental industry. *The American Economic Review*, 102(2):686–719.

Iaria, A. and Wang, A. (2019). Inferring complementarity from correlations rather than structural estimation. *Working Paper*, pages 1–17.

il Kim, K. (2014). Identification of the distribution of random coefficients in static and dynamic discrete choice models. *The Korean Economic Review*, 30(2):191–216.

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

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

Kretschmer, T., Miravete, E. J., and Pernías, J. C. (2012). Competitive pressure and the adoption of complementary innovations. *American Economic Review*, 102(4):1540–70.

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

Lee, S., Kim, J., and Allenby, G. M. (2013). A direct utility model for asymmetric complements. *Marketing Science*, 32(3):454–470.

Lewbel, A. (1985). Bundling of substitutes or complements. *International Journal of Industrial Organization*, 3(1):101–107.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics*, 30(1):67–80.

Lewbel, A. (2019). The identification zoo: Meanings of identification in econometrics. *Journal of Economic Literature*, 57(4):835–903.

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

Liu, H., Chintagunta, P. K., and Zhu, T. (2010). Complementarities and the demand for home broadband internet services. *Marketing Science*, 29(4):701–720.

Manski, C. F. (1989). Anatomy of the selection problem. *Journal of Human resources*, pages 343–360.

Manski, C. F. (2003). *Partial identification of probability distributions*. Springer Science & Business Media.

Manski, C. F. and Sherman, L. (1980). An empirical analysis of household choice among motor vehicles. *Transportation Research Part A: General*, 14(5-6):349–366.

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

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

McAfee, R. P., McMillan, J., and Whinston, M. D. (1989). Multiproduct monopoly, commodity bundling, and correlation of values. *The Quarterly Journal of Economics*, 104(2):371–383.

McFadden, D. and Train, K. (2000). Mixed mnl models for discrete response. *Journal of applied Econometrics*, pages 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*, pages 395–421.

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

Nevo, A., Rubinfeld, D. L., and McCabe, M. (2005). Academic journal pricing and the demand of libraries. *American Economic Review*, 95(2):447–452.

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

Pozzi, A. (2012). Shopping cost and brand exploration in online grocery. *American Economic Journal: Microeconomics*, 4(3):96–120.

Romano, J. P. and Shaikh, A. M. (2012). On the uniform asymptotic validity of subsampling and the bootstrap. *The Annals of Statistics*, 40(6):2798–2822.

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

Ruiz, F. J., Athey, S., and Blei, D. M. (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*, pages 1605–1633.

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

Song, M., Nicholson, S., and Lucarelli, C. (2017). Mergers with interfirm bundling: a case of pharmaceutical cocktails. *The RAND Journal of Economics*, 48(3):810–834.

Thomassen, Ø., Smith, H., Seiler, S., and Schiraldi, P. (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.*

Wang, A. (2019). A blp demand model of product-level market shares with complementarity. *Working Paper.*

Zhou, J. (2017). Competitive bundling. *Econometrica*, 85(1):145–172.

8 Appendix

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

In this Appendix, we illustrate that the model of preference for variety originally proposed by Hendel (1999) in the context of demand for computers and then applied by Dubé (2004) in the context of demand for soft drinks is a special case of model (3). In particular, Hendel (1999)'s model is a version of model (3) in which each demand synergy parameter $\Gamma_{it\mathbf{b}}$ is restricted to be negative 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 choice 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. 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 to the 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.[34] 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 purchasing bundle $\mathbf{b} = ((j_1, q_1), ..., (j_{K_i}, q_{K_i})) \in \mathbf{C}_i$ is:

$$
\begin{aligned}
U_{i\mathbf{b}} &= \sum_{k=1}^{K_i} (\Psi_{ij_k k} q_k)^\alpha S_i - \sum_{k=1}^{K_i} p_{j_k} q_k + y_i \\
&= \sum_{k=1}^{K_i} (\Psi_{ij_k k} q_k)^\alpha S_i + \sum_{k=1}^{K_i} (\Psi_{ij_k k})^\alpha S_i q_k - \sum_{k=1}^{K_i} (\Psi_{ij_k k})^\alpha S_i q_k - \sum_{k=1}^{K_i} p_{j_k} q_k + y_i \\
&= \sum_{k=1}^{K_i} (\Psi_{ij_k k} q_k)^\alpha S_i + \sum_{k=1}^{K_i} \sum_{q=1}^{q_k} (\Psi_{ij_k k})^\alpha S_i - \sum_{k=1}^{K_i} (\Psi_{ij_k k})^\alpha S_i q_k - \sum_{k=1}^{K_i} \sum_{q=1}^{q_k} p_{j_k} + y_i \\
&= \sum_{k=1}^{K_i} \sum_{q=1}^{q_k} [(\Psi_{ij_k k})^\alpha S_i - p_{j_k}] + \sum_{k=1}^{K_i} (\Psi_{ijk})^\alpha S_i [q_k^\alpha - q_k] + y_i \\
&= \sum_{k=1}^{K_i} \sum_{q=1}^{q_k} u_{ij_k k} + \Gamma_{i\mathbf{b}} + y_i,
\end{aligned}
\tag{19}
$$

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 (19) 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 i.i.d. Gumbel error terms.

8.2 Proof of Lemma 1

To prove the first statement, we show that given a distribution function for β_{it}, $F(\cdot; \Sigma'_F)$, there exists a unique $\delta'_t \in \mathbb{R}^{C_{t1}}$ for $t = 1, ..., T$ that solves $s_t(\delta'_t; \Sigma'_F) = \mathcal{s}_t$. This is equivalent to showing that given $F(\cdot; \Sigma'_F)$, the market share function $s_t(\cdot; \Sigma'_F)$ is invertible for $t = 1, ..., T$. Because our arguments with regard to the first statement do not depend on whether F is parametric or non-parametric, hereafter we denote $F(\cdot; \Sigma'_F)$ simply 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:

$$
\mathbb{J}_t(\delta'_t; F) = \frac{\partial s_t}{\partial \delta'_t}(\delta'_t; F) = \left(\frac{\partial s_{t\mathbf{b}}}{\partial \delta'_{t\mathbf{b}'}}(\delta'_t; F) \right)_{\mathbf{b}, \mathbf{b}' \in \mathbf{C}_{t1}}. \tag{20}
$$

[34] 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 $q_{j_k k} = q_k$.

Corollary 2 from Berry et al. (2013) provides sufficient conditions for the invertibility of differentiable market share functions. We now verify that market share function (3) satisfies the two sufficient conditions of Corollary 2 from Berry et al. (2013): (a) weak substitutes (Assumption 2 in Berry et al. (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 $\mathbf{b}, \mathbf{b'} \in \mathbf{C}_{t1}, \mathbf{b} \neq \mathbf{b'}$:

$$\begin{aligned} \frac{\partial s_{t\mathbf{b}}}{\partial \delta'_{t\mathbf{b}}}(\delta'_t; F) &= \int s_{it\mathbf{b}}(\delta'_t; \beta_{it})(1 - s_{it\mathbf{b}}(\delta'_t; \beta_{it}))dF(\beta_{it}) \\ \frac{\partial s_{t\mathbf{b}}}{\partial \delta'_{t\mathbf{b'}}}(\delta'_t; F) &= -\int s_{it\mathbf{b}}(\delta'_t; \beta_{it}) s_{it\mathbf{b'}}(\delta'_t; \beta_{it})dF(\beta_{it}). \end{aligned} \tag{21}$$

As discussed by Berry et al. (2013), the weak substitutes condition does not rule out complementarity in a discrete choice model in which alternatives are defined as bundles, as in demand model (3). 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 (21). 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 (21), we can re-write $\mathbb{J}_t(\delta'_t; F)$ as:

$$\mathbb{J}_t(\delta'_t; F) = \int \left[\mathrm{Diag}(s_{it}(\delta'_t; \beta_{it})) - s_{it}(\delta'_t; \beta_{it}) s_{it}(\delta'_t; \beta_{it})^{\mathrm{T}}\right] dF(\beta_{it}), \tag{22}$$

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:

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

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 corresponding (non-degenerate) eigenvector by x. Without loss of generality, suppose that the maximal element of vector 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:

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

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 et al. (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.

We now prove the second statement of the Lemma. According to Assumption 1, the density function $\frac{dF(\beta_{it}; \Sigma_F')}{d\beta_{it}}$ is continuously differentiable with respect to Σ_F'. As a consequence, $s_t(\delta_t'; \Sigma_F') - \mathscr{s}_t'$ is continuously differentiable with respect to $(\delta_t', \mathscr{s}_t', \Sigma_F')$. As we showed above, the Jacobian matrix $\frac{\partial [s_t(\delta_t'; \Sigma_F') - \mathscr{s}_t']}{\partial \delta_t'}\Big|_{(\delta_t', \mathscr{s}_t', \Sigma_F') = (\delta_t, \mathscr{s}_t, \Sigma_F)} = \mathbb{J}_t(\delta_t; F(\cdot; \Sigma_F))$ is invertible. Then, according to the Implicit Function Theorem, in a neighbourhood of $(\delta_t, \mathscr{s}_t, \Sigma_F)$, for any $(\mathscr{s}_t', \Sigma_F')$ there exists a unique δ_t' such that $s_t(\delta_t'; \Sigma_F') = \mathscr{s}_t'$ and $s_t^{-1}(\mathscr{s}_t'; \Sigma_F') = \delta_t'$ is continuously differentiable with respect to $(\mathscr{s}_t', \Sigma_F')$. This completes the proof of the second statement.

8.3 Proof of Rank Regularity Property

Without loss of generality, suppose that $\Theta_\Sigma \subset \Upsilon$ is a compact set, where $\Upsilon \subset \mathbb{R}^{P+D}$ is a topological space of $\mathbb{R}^{P+D}$. Moreover, according to Assumption 1, $\frac{\partial m(\Sigma')}{\partial \Sigma'}$ 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.

8.4 Preliminaries for Theorem 1

Here we report a preliminary Lemma useful to prove Theorem 1.

Lemma 2. *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 and the differentiability of $g(\Sigma_g)$ with respect to Σ_g in Assumption 2. 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 the first-order Taylor expansion, we have:

$$\begin{aligned} m(\Sigma_N) &= m(\Sigma) + \frac{\partial m(\Sigma')}{\partial \Sigma'}\bigg|_{\Sigma'=\Sigma} (\Sigma_N - \Sigma) + o(|\Sigma_N - \Sigma|), \\ \frac{\partial m(\Sigma')}{\partial \Sigma'}\bigg|_{\Sigma'=\Sigma} \frac{\Sigma_N - \Sigma}{|\Sigma_N - \Sigma|} &= -\frac{o(|\Sigma_N - \Sigma|)}{|\Sigma_N - \Sigma|}, \end{aligned} \tag{23}$$

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 $\mathbb{R}^{P+D}$, which is compact. Then, there exists a subsequence $\left\{\frac{\Sigma_{N_\ell} - \Sigma}{|\Sigma_{N_\ell} - \Sigma|}\right\}$ and $v \in \mathbb{R}^{P+D}$ with $|v| = 1$, such that $\frac{\Sigma_{N_\ell} - \Sigma}{|\Sigma_{N_\ell} - \Sigma|} \to v$. By applying the second equation of (23) 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 neighbourhood 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 $x \in \mathbb{R}^{P+D}$ 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 $m(\Sigma') = 0$. □

8.5 Proof of Theorem 1

Sufficiency. We prove sufficiency 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:

$$\begin{aligned}
&\delta^N_{t\mathbf{b}}(\Gamma^N_{\mathbf{b}}) + g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma^N_g) = s^{-1}_{t\mathbf{b}}(\mathcal{s}_t; \Sigma^N_F),\\
&\delta^N_{tj} = s^{-1}_{tj}(\mathcal{s}_t; \Sigma^N_F),\ j \in \mathbf{b},\\
&\Gamma^N_{\mathbf{b}} = s^{-1}_{t\mathbf{b}}(\mathcal{s}_t; \Sigma^N_F) - \sum_{j\in\mathbf{b}} s^{-1}_{tj}(\mathcal{s}_t; \Sigma^N_F) - g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma^N_g).
\end{aligned} \tag{24}$$

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 2, then 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 (24), then for all $N \geq N_0$, $\Gamma^N_{\mathbf{b}} = \Gamma_{\mathbf{b}}$. Moreover, because of the first two equations of (24), we have $\delta^N_{t\mathbf{b}} = \delta_{t\mathbf{b}}$, for all $N \geq N_0$, $t = 1, ..., T$ and $\mathbf{b} \in \mathbf{C}_{t1}$. As a consequence, $(\delta^N_{1\mathbf{J}_1}, ..., \delta^N_{T\mathbf{J}_T}, \Gamma^N, \Sigma^N) = (\delta_{1\mathbf{J}_1}, ..., \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$ for all $N \geq N_0$, which contradicts $(\delta^N_{1\mathbf{J}_1}, ..., \delta^N_{T\mathbf{J}_T}, \Gamma^N, \Sigma^N) \neq (\delta_{1\mathbf{J}_1}, ..., \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$ for any N.

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 the definition of rank regularity in footnote 11, there exists a neighbourhood of the true Σ, U, such that $\text{rank}(\frac{m(\Sigma')}{\partial \Sigma'}) = \text{rank}(\frac{m(\Sigma')}{\partial \Sigma'})\big|_{\Sigma'=\Sigma} = r$ for each $\Sigma' \in U$. By applying the Constant Rank Theorem at $\Sigma' = \Sigma$, there are open sets $U_1, U_2 \subset \mathbb{R}^{P+D}$ and $U_3 \subset \mathbb{R}^Q$ and diffeomorphisms $\phi : U_1 \to U_2$, $\psi : U_3 \to U_3$ such that $\Sigma \in U_1 \subset U$ and $\psi \circ m \circ \phi^{-1}(x') = (x'_1, \dots, x'_r, 0, \dots, 0)$ for all $x' \in U_2$.[35]

Define $x = (x_1, ..., x_{P+D}) = \phi(\Sigma)$ and a sequence $\{x^N = (x^N_1, ..., x^N_{P+D})\}$ such that $x^N_\ell = x_\ell$, for $\ell = 1, ..., r$ and $x^N_\ell = x_\ell + \frac{1}{N}$, for N large enough so that $x^N \neq x$ and $x^N \in U_2$. Note that

$$\begin{aligned}
\psi \circ m \circ \phi^{-1}(x) &= (x_1, ...x_r, 0, \dots, 0)\\
&= (x^N_1, ...x^N_r, 0, \dots, 0)\\
&= \psi \circ m \circ \phi^{-1}(x^N)
\end{aligned} \tag{25}$$

and that

$$\begin{aligned}
\psi \circ m \circ \phi^{-1}(x) &= \psi \circ m \circ \phi^{-1}(\phi(\Sigma))\\
&= \psi \circ m(\Sigma)\\
&= \psi(0).
\end{aligned} \tag{26}$$

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

[35] For the details of the Constant Rank Theorem, see Theorem 7.1 by Boothby (1986).

$m(\Sigma^N) = 0$ for each N. According to (24) from the proof of sufficiency, 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 and this concludes the proof.

8.6 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 a 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 1, model (3) is not locally identified and this concludes the proof.

8.7 Testing Procedures for Assumption 2 and Assumption 3

In this section, we develop testing procedures for Assumptions 2 and 3 on the basis of partial identification methods. For a given subset of markets $\mathbf{T}_0 \subset \mathbf{T}$, the identification set of $\theta = ((\delta_{t\mathbf{J}_t})_{t\in\mathbf{T}_0}, \Gamma, \Sigma_F, \Sigma_g)$ is defined by the moment equalities:

$$s_{t\mathbf{b}}(\delta_t(\Gamma + g_t(\Sigma_g)); \Sigma_F) = \mathcal{s}_{t\mathbf{b}}, \tag{27}$$

for $t \in \mathbf{T}_0$ and $\mathbf{b} \in \mathbf{C}_{t1}$, where $g_t(\Sigma_g) = (g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g))_{\mathbf{b}\in\mathbf{C}_{t2}}$ and $\delta_t(\Gamma+g_t(\Sigma_g)) = (\delta_{t1}, ..., \delta_{tJ_t}, (\delta_{t\mathbf{b}}(\Gamma_\mathbf{b}+g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g)))_{\mathbf{b}\in\mathbf{C}_{t2}})$. We denote by $\Theta(\mathbf{T}_0)$ the identification set of θ defined by (27) and by $Q((\mathcal{s}_t)_{t\in\mathbf{T}_0}, \theta')$ the following criterion function:

$$Q((\mathcal{s}_t)_{t\in\mathbf{T}_0}, \theta') = \sum_{t\in\mathbf{T}_0}(s_t(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F) - \mathcal{s}_t)^{\mathrm{T}}\Omega_t^{-1}(s_t(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F) - \mathcal{s}_t). \tag{28}$$

Note that $Q((\mathcal{s}_t)_{t\in\mathbf{T}_0}, \theta') = 0$ if and only if $\theta' \in \Theta(\mathbf{T}_0)$. Denote by $I_{t\mathbf{b}}$ the number of individuals in market t observed to choose $\mathbf{b}$ and by $\hat{\mathcal{s}}_{t\mathbf{b}} = \frac{I_{t\mathbf{b}}}{I}$ the corresponding observed market share. As I increases to infinity, $\hat{\mathcal{s}}_{t\mathbf{b}} \xrightarrow{p} \mathcal{s}_{t\mathbf{b}}$ and $\sqrt{I}(\hat{\mathcal{s}}_t - \mathcal{s}_t) \xrightarrow{p} \mathcal{N}(0, \Omega_t)$ for $t = 1, ..., T$, $\mathbf{b} \in \mathbf{C}_t$, where $\Omega_t = (\omega_{t\mathbf{bb}'})_{\mathbf{b},\mathbf{b}'\in\mathbf{C}_{t1}}$ with $\omega_{t\mathbf{bb}'} = \mathcal{s}_{t\mathbf{b}}(1 - \mathcal{s}_{t\mathbf{b}})$ when $\mathbf{b} = \mathbf{b}'$ and $\omega_{t\mathbf{bb}'} = -\mathcal{s}_{t\mathbf{b}}\mathcal{s}_{t\mathbf{b}'}$ otherwise.

Denote by $\hat{\Omega}_t$ an estimator of Ω_t that satisfies $\hat{\Omega}_t \xrightarrow{p} \Omega_t$ and $\sqrt{I}(\hat{\Omega}_t - \Omega_t) = O_p(1)$.[36] We then define the sample counterpart of criterion function $Q(\cdot)$ as:

$$Q_I((\hat{\jmath}_t)_{t\in\mathbf{T}_0}, \theta') = \sum_{t\in\mathbf{T}_0} (s_t(\delta_t'(\Gamma' + g_t(\Sigma_g')); \Sigma_F') - \hat{\jmath}_t)^{\mathrm{T}} \hat{\Omega}_t^{-1} (s_t(\delta_t'(\Gamma' + g_t(\Sigma_g')); \Sigma_F') - \hat{\jmath}_t). \quad (29)$$

Testing Procedure for Assumption 2. In this section, we maintain $\mathbf{T}_0 = \mathbf{T}$. Note that Assumption 2 holds if and only if $\Theta(\mathbf{T}) \neq \emptyset$, i.e., there is at least a profile of θ that satisfies moment equalities (27). Hence, we propose a specification test on the basis of the following hypotheses:

$$\mathrm{H}_0 : \Theta(\mathbf{T}) \neq \emptyset \text{ versus } \mathrm{H}_1 : \Theta(\mathbf{T}) = \emptyset. \quad (30)$$

Denote by $q^{1-\alpha}_{\sum_{t\in\mathbf{T}} C_{t1}}$ the $1-\alpha$ quantile of $\chi^2\left(\sum_{t\in\mathbf{T}} C_{t1}\right)$ and define the following random set:

$$\Theta_I(\mathbf{T}) = \{\theta' \in \Theta : I \cdot Q_I((\hat{\jmath}_t)_{t\in\mathbf{T}}, \theta') \leq q^{1-\alpha}_{\sum_{t\in\mathbf{T}} C_{t1}}\}. \quad (31)$$

If $\Theta_I(\mathbf{T}) = \emptyset$, then we reject H_0 from (30).

Proposition 1. *Under H_0 from* (30), $\limsup_{I\to\infty} \sup_{\theta'\in\Theta(\mathbf{T})} \Pr\left[\Theta_I(\mathbf{T}) = \emptyset\right] \leq \alpha$.

Proof. Under H_0 from (30), for any $\theta' \in \Theta(\mathbf{T})$, we have:

$$\begin{aligned} \Pr\left[\Theta_I(\mathbf{T}) = \emptyset\right] &\leq \Pr\left[\theta' \notin \Theta_I(\mathbf{T})\right] \\ &= \Pr\left[I \cdot Q_I((\hat{\jmath}_t)_{t\in\mathbf{T}}, \theta') > q^{1-\alpha}_{\sum_{t\in\mathbf{T}} C_{t1}}\right] \\ &= \Pr\left[\sum_{t\in\mathbf{T}} \left[\sqrt{I}(\jmath_t - \hat{\jmath}_t)\right]^{\mathrm{T}} \hat{\Omega}_t^{-1} \left[\sqrt{I}(\jmath_t - \hat{\jmath}_t)\right] > q^{1-\alpha}_{\sum_{t\in\mathbf{T}} C_{t1}}\right]. \end{aligned} \quad (32)$$

Since $\hat{\Omega}_t \xrightarrow{p} \Omega_t$ (and hence $\hat{\Omega}_t^{-1} \xrightarrow{p} \Omega_t^{-1}$) and $\sqrt{I}(\hat{\jmath}_t - \jmath_t) \xrightarrow{p} \mathcal{N}(0, \Omega_t)$, for $t \in \mathbf{T}$, we obtain:

$$\sum_{t\in\mathbf{T}} \left[\sqrt{I}(\jmath_t - \hat{\jmath}_t)\right]^{\mathrm{T}} \hat{\Omega}_t^{-1} \left[\sqrt{I}(\jmath_t - \hat{\jmath}_t)\right] \xrightarrow{d} \chi^2\left(\sum_{t\in\mathbf{T}} C_{t1}\right).$$

Note that the probability on the right-hand side of (32) converges to α and does not depend on θ'. Then,

$$\limsup_{I\to\infty} \sup_{\theta'\in\Theta(\mathbf{T})} \Pr\left[\Theta_I(\mathbf{T}) = \emptyset\right] \leq \alpha.$$

The proof is completed. □

[36] Such an estimator can be $\hat{\Omega}_t = (\hat{\omega}_{t\mathbf{b}\mathbf{b}'})_{\mathbf{b},\mathbf{b}'\in\mathbf{C}_{t1}}$, where $\hat{\omega}_{t\mathbf{b}\mathbf{b}'} = \hat{\jmath}_{t\mathbf{b}}(1-\hat{\jmath}_{t\mathbf{b}})$ when $\mathbf{b} = \mathbf{b}'$ and $\hat{\omega}_{t\mathbf{b}\mathbf{b}'} = -\hat{\jmath}_{t\mathbf{b}}\hat{\jmath}_{t\mathbf{b}'}$ otherwise.

Testing Procedure for Assumption 3. In this section, we assume that $\Theta(\mathbf{T}) \neq \emptyset$ and therefore that $\Theta(\mathbf{T}_0) \neq \emptyset$ for any $\mathbf{T}_0 \subset \mathbf{T}$. Given $\mathbf{T}_0$, we first derive a consistent estimator for $\Theta(\mathbf{T}_0)$:

Lemma 3. *Define a sequence of random sets:*

$$\mathscr{C}(a_I) = \{\theta' \in \Theta : I \cdot Q_I((\hat{\jmath}_t)_{t\in\mathbf{T}_0}, \theta') \leq a_I\},$$

where $a_I \geq 0$ satisfies $a_I \to \infty$ and $\frac{a_I}{I} \to 0$. Then,

$$\lim_{I\to\infty} \inf_{\theta'\in\Theta(\mathbf{T}_0)} \Pr\left[\Theta(\mathbf{T}_0) \subset \mathscr{C}(a_I)\right] = 1$$

and

$$\lim_{I\to\infty} \sup_{\theta'\in\Theta(\mathbf{T}_0)} d_H(\Theta(\mathbf{T}_0), \mathscr{C}(a_I)) = 0,$$

where $d_H(\cdot,\cdot)$ is the Hausdorff metric: $d_H(A,B) = \sup_{a\in A}\inf_{b\in B}|a-b| + \sup_{b\in B}\inf_{a\in A}|b-a|$.

Proof. See Theorem 3.1 by Chernozhukov et al. (2007). □

The choice of a_I is up to the econometrician. For example, one can choose $a_I = \ln I$ (see Chernozhukov et al. (2007) for a detailed discussion). In what follows, we focus on situations in which $\Theta(\mathbf{T}_0)$ contains only interior points of Θ. While Assumption 3 is abstract and not easy to test directly, we propose the following Condition and show that it implies Assumption 3:

Condition 1. *There exists $\mathbf{T}_0 \subsetneq \mathbf{T}$ such that $\frac{\partial m(\Sigma';\mathbf{T}_0)}{\partial \Sigma'}$ is of full column rank when evaluated at any of the solutions to $m(\Sigma';\mathbf{T}_0) = 0$.*

Remark 4. *Denote by $\Theta_\Sigma(\mathbf{T}_0)$ the set of solutions to $m(\Sigma';\mathbf{T}_0) = 0$. Since the true parameters $\Sigma_0 \in \Theta_\Sigma(\mathbf{T}_0)$, Condition 1 implies that $\frac{\partial m(\Sigma';\mathbf{T}_0)}{\partial \Sigma'}$ is of full column rank when evaluated at $\Sigma' = \Sigma_0$. As a consequence, model* (3) *is locally identified according to Theorem 1.*

Remark 5. *As shown in section 3, $\theta' \in \Theta(\mathbf{T}_0)$ holds if and only if $m(\Sigma';\mathbf{T}_0) = 0$. Then, $\Theta_\Sigma(\mathbf{T}_0)$ is the projection of $\Theta(\mathbf{T}_0)$ along the dimensions of Σ. Moreover, because of Lemma 3, the projection of $\mathscr{C}(a_n)$ along the dimensions of Σ also defines a consistent estimator for $\Theta_\Sigma(\mathbf{T}_0)$ which covers asymptotically $\Theta_\Sigma(\mathbf{T}_0)$ with probability* 1 *and that we denote by $\mathscr{C}_\Sigma(a_n)$.*

The next Proposition shows that Condition 1 is sufficient for Assumption 3:

Proposition 2. *If Condition 1 holds, then Assumption 3 holds.*

Proof. We prove this by contradiction. Denote the solution set of $m(\Sigma';\mathbf{T}_0) = 0$ in Θ_Σ by S. Suppose that S contains infinitely many elements. Because S is a closed subset of the compact set Θ_Σ, S is itself compact. Consequently, because S has infinitely many elements, then there exists an accumulation point $\Sigma_0' \in S$: in any neighbourhood of Σ_0', we can find another $\Sigma_0'' \in S$, i.e. another solution to $m(\Sigma'; \mathbf{T}_0) = 0$. Due to Assumption 3, we know that at $\Sigma_0' \in S$, the corresponding Jacobian matrix $\frac{\partial m(\Sigma'; \mathbf{T}_0)}{\partial \Sigma'}_{\Sigma'=\Sigma_0'}$ is of full column rank. Then, locally, $\Sigma' = \Sigma_0'$ must be the unique solution to $m(\Sigma'; \mathbf{T}_0) = 0$. This contradicts Σ_0' being an accumulation point in S. $\square$

We then propose a test for Assumption 3 on the basis of the following hypotheses:

$$\mathrm{H}_0: \text{ Condition 1 does not hold. versus } \mathrm{H}_1: \text{ Condition 1 holds.} \tag{33}$$

H_0 from (33) is equivalent to the hypothesis that there exists some $\theta' \in \Theta(\mathbf{T}_0)$ such that $\frac{\partial m(\Sigma'; \mathbf{T}_0)}{\partial \Sigma'}$ is rank deficient when evaluated at $\Sigma' = (\Sigma_F', \Sigma_g')$. Define the following function:

Assumption 6. *Suppose that* $J((\jmath_t')_{t\in\mathbf{T}_0}, \Sigma') : \times_{t\in\mathbf{T}_0} \mathcal{S}_t \times \Theta_\Sigma \to \mathbb{R}_+ \cup \{0\}$ *satisfies:*

- $J \geq 0$.
- $J = 0$ *if and only if* $\frac{\partial m(\Sigma'; \mathbf{T}_0)}{\partial \Sigma'}$ *is rank deficient.*

where $\mathcal{S}_t = \{\jmath_t' \in \mathbb{R}^{C_{t1}} : \jmath_{t\mathbf{b}}' > 0$ *and* $\sum_{\mathbf{b}\in\mathbf{C}_{t1}} \jmath_{t\mathbf{b}}' < 1,\ \mathbf{b} \in \mathbf{C}_{t1}.\}$.

Example 1. *The determinant function*

$$J((\jmath_t')_{t\in\mathbf{T}_0}, \Sigma') = Det\left(\left[\left(\frac{\partial m(\Sigma'; \mathbf{T}_0)}{\partial \Sigma'}\right)^T \left(\frac{\partial m(\Sigma'; \mathbf{T}_0)}{\partial \Sigma'}\right)\right]\right)$$

Example 2. *The minimal eigenvalue function*

$$J((\jmath_t')_{t\in\mathbf{T}_0}, \Sigma') = \inf_{|\lambda|=1} \lambda^T \left[\left(\frac{\partial m(\Sigma'; \mathbf{T}_0)}{\partial \Sigma'}\right)^T \left(\frac{\partial m(\Sigma'; \mathbf{T}_0)}{\partial \Sigma'}\right)\right] \lambda,$$

where λ *is unit vector of dimension* $P + D$.

Define the criterion function $J^*((\jmath_t')_{t\in\mathbf{T}_0}) = \inf_{\Sigma'\in\Theta_\Sigma(\mathbf{T}_0)} J((\jmath_t')_{t\in\mathbf{T}_0}, \Sigma')$. Note that $J^*((\jmath_t)_{t\in\mathbf{T}_0}) = 0$ if and only if H_0 from (33) holds. We then propose the following test statistic:

$$J_I^*((\hat{\jmath}_t)_{t\in\mathbf{T}_0}) = \inf_{\Sigma'\in\mathscr{C}_\Sigma(a_n)} J((\hat{\jmath}_t)_{t\in\mathbf{T}_0}, \Sigma') \tag{34}$$

and the next two Theorems establish its properties.

Theorem 7. *Suppose Assumptions 1, 2 and 6 hold. Moreover,* $\sqrt{I}(\hat{\jmath}_t - \jmath_t) \xrightarrow{d} \mathcal{N}(0, \Omega_t)$ *for* $t \in \mathbf{T}$.

- *If J is continuous in $\times_{t\in\mathbf{T}_0}\mathscr{S}_t \times \Theta_\Sigma$, then $J_I^*((\hat{\mathscr{s}}_t)_{t\in\mathbf{T}_0}) \xrightarrow{p} J^*((\mathscr{s}_t)_{t\in\mathbf{T}_0})$, uniformly for $\theta \in \Theta(\mathbf{T}_0)$.*
- *If J is Lipschitz continuous in $\times_{t\in\mathbf{T}_0}\mathscr{S}_t \times \Theta_\Sigma$, then under H_0 from* (33), *$\sqrt{I}\cdot J_I^*((\hat{\mathscr{s}}_t)_{t\in\mathbf{T}_0}) = O_p(1)$.*

Proof. For any $\theta \in \Theta(\mathbf{T})$, on the stochastic event $\{\Theta(\mathbf{T}_0) \subset \mathscr{C}(a_I)\}$, we can write:

$$\begin{aligned} J^*((\mathscr{s}_t)_{t\in\mathbf{T}_0}) &= J((\mathscr{s}_t)_{t\in\mathbf{T}_0}, \Sigma^*), \\ J_I^*((\hat{\mathscr{s}}_t)_{t\in\mathbf{T}_0}) &= J((\hat{\mathscr{s}}_t)_{t\in\mathbf{T}_0}, \Sigma_I^*), \end{aligned} \tag{35}$$

where $\Sigma^* \in \underset{\Sigma'\in\Theta_\Sigma(\mathbf{T}_0)}{\operatorname{argmin}} J((\mathscr{s}_t)_{t\in\mathbf{T}_0}, \Sigma')$ and $\Sigma_I^* \in \underset{\Sigma'\in\mathscr{C}_\Sigma(a_I)}{\operatorname{argmin}} J((\hat{\mathscr{s}}_t)_{t\in\mathbf{T}_0}, \Sigma')$. Note that $\Sigma^* \in \Theta(\mathbf{T}_0) \subset \mathscr{C}_\Sigma\Sigma(a_I)$. Then, we have:

$$J_I^*((\hat{\mathscr{s}}_t)_{t\in\mathbf{T}_0}) - J^*((\mathscr{s}_t)_{t\in\mathbf{T}_0}) \leq J((\hat{\mathscr{s}}_t)_{t\in\mathbf{T}_0}, \Sigma^*) - J((\mathscr{s}_t)_{t\in\mathbf{T}_0}, \Sigma^*). \tag{36}$$

By construction, $\Theta_\Sigma(\mathbf{T}_0)$ is a compact set. Then, there exists $\tilde{\Sigma}_I^* \in \Theta_\Sigma(\mathbf{T}_0)$ such that $d(\Sigma_I^*, \Theta_\Sigma(\mathbf{T}_0)) = d(\Sigma_I^*, \tilde{\Sigma}_I^*)$. Hence, we obtain:

$$\begin{aligned} J_I^*((\hat{\mathscr{s}}_t)_{t\in\mathbf{T}_0}) - J^*((\mathscr{s}_t)_{t\in\mathbf{T}_0}) &= \left[J((\hat{\mathscr{s}}_t)_{t\in\mathbf{T}_0}, \Sigma_I^*) - J((\hat{\mathscr{s}}_t)_{t\in\mathbf{T}_0}, \tilde{\Sigma}_I^*)\right] + \left[J((\hat{\mathscr{s}}_t)_{t\in\mathbf{T}_0}, \tilde{\Sigma}_I^*) - J((\mathscr{s}_t)_{t\in\mathbf{T}_0}, \tilde{\Sigma}_I^*)\right] \\ &\quad + \left[J((\mathscr{s}_t)_{t\in\mathbf{T}_0}, \tilde{\Sigma}_I^*) - J((\mathscr{s}_t)_{t\in\mathbf{T}_0}, \Sigma^*)\right] \\ &\geq \left[J((\hat{\mathscr{s}}_t)_{t\in\mathbf{T}_0}, \Sigma_I^*) - J((\hat{\mathscr{s}}_t)_{t\in\mathbf{T}_0}, \tilde{\Sigma}_I^*)\right] + \left[J((\hat{\mathscr{s}}_t)_{t\in\mathbf{T}_0}, \tilde{\Sigma}_I^*) - J((\mathscr{s}_t)_{t\in\mathbf{T}_0}, \tilde{\Sigma}_I^*)\right]. \end{aligned} \tag{37}$$

According to Lemma 3, $d_H(\Theta(\mathbf{T}_0), \mathscr{C}(a_I)) \to 0$ uniformly for $\theta \in \Theta(\mathbf{T}_0)$. We then obtain $d(\Sigma_I^*, \tilde{\Sigma}_I^*) \to 0$ uniformly for $\theta \in \Theta(\mathbf{T}_0)$.

Suppose that J is continuous in $\times_{t\in\mathbf{T}_0}\mathscr{S}_t \times \Theta_\Sigma$. Then, in a compact set $\mathscr{S}^* \times \Theta$, where $\mathscr{S}^*$ is a compact neighbourhood of $(\mathscr{s}_t)_{t\in\mathbf{T}_0}$, J is uniformly continuous. Together with $d(\Sigma_I^*, \tilde{\Sigma}_I^*) \to 0$ uniformly for $\theta \in \Theta(\mathbf{T}_0)$, we obtain that the right-hand side of (36) and that of (37) converge to 0 on $\{\Theta(\mathbf{T}_0) \subset \mathscr{C}(a_I)\}$, uniformly for $\theta \in \Theta(\mathbf{T}_0)$. Note that $\{\Theta(\mathbf{T}_0) \subset \mathscr{C}(a_I)\}$ holds asymptotically with probability 1, uniformly for $\theta \in \Theta(\mathbf{T}_0)$. This proves the first statement.

Suppose that J is Lipschitz continuous in $\times_{t\in\mathbf{T}_0}\mathscr{S}_t \times \Theta_\Sigma$. Under H_0 from (33), we have $J^*((\mathscr{s}_t)_{t\in\mathbf{T}_0}) = 0$. Then, by applying the Mean Value Theorem on the right-hand side of (36), we obtain that on $\{\Theta(\mathbf{T}_0) \subset \mathscr{C}(a_I)\}$:

$$\begin{aligned} 0 \leq J_I^*((\hat{\mathscr{s}}_t)_{t\in\mathbf{T}_0}) &= J_I^*((\hat{\mathscr{s}}_t)_{t\in\mathbf{T}_0}) - J^*((\mathscr{s}_t)_{t\in\mathbf{T}_0}) \\ &\leq J((\hat{\mathscr{s}}_t)_{t\in\mathbf{T}_0}, \Sigma^*) - J((\mathscr{s}_t)_{t\in\mathbf{T}_0}, \Sigma^*) \leq L|\hat{\mathscr{s}} - \mathscr{s}|, \end{aligned} \tag{38}$$

where L is the Lipschitz constant of $J(\cdot)$. Then, by using $\sqrt{I}(\hat{\mathscr{s}}_t - \mathscr{s}_t) \xrightarrow{d} \mathscr{N}(0, \Omega_t)$, we obtain that:

$$0 \leq \sqrt{I} \cdot J_I^*((\hat{\jmath}_t)_{t\in\mathbf{T}_0}) \leq L|\sqrt{I}(\hat{\jmath} - \jmath)| = O_p(1) \tag{39}$$

and the second statement is proved. □

We now illustrate how to approximate the quantiles of $J_I^*((\hat{\jmath}_t)_{t\in\mathbf{T}_0})$ under H_0 from (33) by bootstrap methods building on Romano and Shaikh (2012). Denote by $\mathbf{P}_R^I = \{P_s^I : s \in \mathbb{R}^R, s_r > 0, \sum_{r=1}^R s_r = 1\}$ the set of multinomial distributions with R outcomes out of I trials. Define the distance ρ on $\mathbf{P}_R^I$ as $\rho(P_s^I, P_{s'}^I) = \sum_{r=1}^R |s_r - s'_r|$ and $\mathbf{P}^I = \times_{t\in\mathbf{T}_0}\mathbf{P}_{C_{t1}}^I$ as the set of joint distributions $P = (P_{s_t}^I)_{t\in\mathbf{T}_0}$, where each $P_{s_t}^I$ is independently distributed across $t \in \mathbf{T}_0$. Note that ρ can be extended to any $P^I = (P_{s_t}^I)_{t\in\mathbf{T}_0}$ and $Q^I = (P_{s'_t}^I)_{t\in\mathbf{T}_0}$ in $\mathbf{P}^I$ as: $\rho(P, Q) = \sum_{t\in\mathbf{T}_0} \rho(P_{s_t}^I, P_{s'_t}^I)$. For any I and any realization $\omega = (i_t)_{t\in\mathbf{T}_0}$ of P, we can define the non-negative random variable $J_I^*(\omega; P) = J_I^*(\left(\frac{i_t}{I}\right)_{t\in\mathbf{T}_0})$. Denote the distribution function of $J_I^*(\omega; P)$ evaluated at $x \geq 0$ by $G_I(x, P)$.

Theorem 8. *Suppose the same Assumptions of Theorem 7 hold. If J is Lipschitz continuous in $\times_{t\in\mathbf{T}_0}\mathcal{S}_t \times \Theta_\Sigma$, then under H_0 from (33), for any α_1, α_2 such that $\alpha_1 \geq 0$, $\alpha_2 \geq 0$, $\alpha_1 + \alpha_2 < 1$,*

$$\liminf_{I\to\infty} \inf_{\theta\in\Theta(\mathbf{T}_0)} \Pr\left[G_I^{-1}(\alpha_1, \hat{P}^I) < \sqrt{I} \cdot J_I^*((\hat{\jmath}_t)_{t\in\mathbf{T}_0}) \leq G_I^{-1}(1 - \alpha_2, \hat{P}^I)\right] \geq 1 - \alpha_1 - \alpha_2,$$

where $\hat{P}^I = (P_{\hat{\jmath}_t}^I)_{t\in\mathbf{T}_0}$.

Proof. Our proof builds on Theorem 2.4 of Romano and Shaikh (2012). It suffices to verify two conditions. First, we show that for any sequences Q^I and P^I in $\mathbf{P}^I$ satisfying $\rho(Q^I, P^I) \to 0$, we have:

$$\lim_{I\to\infty} \sup_{x\geq 0}\{|G_I(x, Q^I) - G_I(x, P^I)|\} \to 0.$$

This can be seen from the construction of $J_I^*(\omega; P)$. For any $x \geq 0$, we have:

$$\begin{aligned} G_I(x, P) &= \Pr\left[J_I^*(\omega; P) \leq x\right] \\ &= \Pr_P\left[(i_t)_{t\in\mathbf{T}_0} : J_I^*\left(\left(\frac{i_t}{I}\right)_{t\in\mathbf{T}_0}\right) \leq x\right] \\ &= \Pr_P\left[(i_t)_{t\in\mathbf{T}_0} : \inf_{\Sigma'\in\mathscr{C}_\Sigma(a_I)} J\left(\left(\frac{i_t}{I}\right)_{t\in\mathbf{T}_0}, \Sigma'\right) \leq x\right] \\ &= \Pr_P\left[(i_t)_{t\in\mathbf{T}_0} : (i_t)_{t\in\mathbf{T}_0} \in \mathscr{N}(x, J(\cdot), a_I, I, \mathbf{T}_0)\right], \end{aligned} \tag{40}$$

where $\mathscr{N}(x, J(\cdot), a_I, I, \mathbf{T}_0)$ is the set of realizations for which $\inf_{\Sigma'\in\mathscr{C}_\Sigma(a_I)} J\left(\left(\frac{i_t}{I}\right)_{t\in\mathbf{T}_0}, \Sigma'\right) \leq x$

holds. Then, given $(x, J(\cdot), a_I, I, \mathbf{T}_0)$, we obtain:

$$\begin{aligned}|G_I(x, Q^I) - G_I(x, P^I)| &= |\mathrm{Pr}_{Q^I}\left[(i_t)_{t\in\mathbf{T}_0} : (i_t)_{t\in\mathbf{T}_0} \in \mathscr{N}(x, J(\cdot), a_I, I, \mathbf{T}_0)\right] \\ &\quad - \mathrm{Pr}_{P^I}\left[(i_t)_{t\in\mathbf{T}_0} : (i_t)_{t\in\mathbf{T}_0} \in \mathscr{N}(x, J(\cdot), a_I, I, \mathbf{T}_0)\right]| \\ &\leq \rho(Q^I, P^I).\end{aligned}$$

Because $\sup_{x\geq 0}\{|G_I(x, Q^I) - G_I(x, P^I)|\} \leq \rho(Q^I, P^I)$ and $\rho(Q^I, P^I) \to 0$, the first condition is verified. We now move on to verifying the second condition. For any sequence $P^I \in \{(P^I_{\jmath_t})_{t\in\mathbf{T}_0} : s_t(\delta_t(\Gamma + g_t(\Sigma_g)); \Sigma_F) = \jmath_t, (\delta_{\mathbf{J}}, \Gamma, \Sigma_F, \Sigma_g) \in \Theta(\mathbf{T}_0)\}$, we have $\rho(\hat{P}^I, P^I) \xrightarrow{p} 0$. This condition holds because $\hat{\jmath}_t$ converges in probability to $\jmath_t$ for any $t \in \mathbf{T}_0$. This completes the proof. $\square$

Finally, for $0 < \alpha < 1$, we propose the following rejection region for test (34):

$$\{\sqrt{I} \cdot J^*_I((\hat{\jmath}_t)_{t\in\mathbf{T}_0}) > G_I^{-1}(1 - \alpha, \hat{P}^I)\}.$$

According to Theorem 7, test (34) has asymptotically unit power, uniformly for $\theta \in \Theta(\mathbf{T}_0)$. Moreover, according to Theorem 8, the size of test (34) is controlled by α, uniformly for $\theta \in \Theta(\mathbf{T}_0)$.

8.8 Proof of Theorem 2

For this result, our arguments do not depend on whether the distribution of random coefficients is parametric or non-parametric and we then denote $F(\cdot; \Sigma_F)$ simply by F. Remember that

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

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

Lemma 4. *For any non-negative integer l,*

$$\sup_{\delta_t, \beta_{it}} \left|\frac{\partial^l s_{it\mathbf{b}}(\delta_t; \beta_{it})}{\partial\delta_{t\mathbf{b}}^l}\right| \leq A_l l!,$$

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

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

$$
\begin{aligned}
e^{\delta_{t\mathbf{b}}+\mu_{it\mathbf{b}}(\beta_{it})} &= s_{it\mathbf{b}} \sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'}+\mu_{it\mathbf{b}'}(\beta_{it})} \\
e^{\delta_{t\mathbf{b}}+\mu_{it\mathbf{b}}(\beta_{it})} &= \frac{\partial^l e^{\delta_{t\mathbf{b}}+\mu_{it\mathbf{b}}(\beta_{it})}}{\partial \delta_{t\mathbf{b}}^l} \\
&= \frac{\partial^l \left(s_{it\mathbf{b}} \sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'}+\mu_{it\mathbf{b}'}(\beta_{it})}\right)}{\partial \delta_{t\mathbf{b}}^l} \\
&= \sum_{k=0}^{l} C_l^k \frac{\partial^k s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^k} \frac{\partial^{l-k} \sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'}+\mu_{it\mathbf{b}'}(\beta_{it})}}{\partial \delta_{t\mathbf{b}}^{l-k}} \\
&= \frac{\partial^l s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^l} \sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'}+\mu_{it\mathbf{b}'}(\beta_{it})} + \sum_{k=0}^{l-1} C_l^k \frac{\partial^k s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^k} e^{\delta_{t\mathbf{b}}+\mu_{it\mathbf{b}}(\beta_{it})}, \\
\frac{\partial^l s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^l} &= s_{it\mathbf{b}} \left(1 - \sum_{k=0}^{l-1} C_l^k \frac{\partial^k s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^k}\right), \\
\left|\frac{\partial^l s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^l}\right| &\leq 1 + \sum_{k=0}^{l-1} C_l^k \left|\frac{\partial^k s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^k}\right|, \\
a_l &\leq 1 + \sum_{k=0}^{l-1} C_l^k a_k, \\
\frac{a_l}{l!} &\leq \frac{1}{l!} + \sum_{k=0}^{l-1} \frac{a_k}{k!} \frac{1}{(l-k)!}.
\end{aligned}
\tag{41}
$$

We now show that $\frac{a_l}{l!} \leq A_l$ by induction. For $l = 0$, the result holds trivially. For $l = 1$, we have $a_1 = \sup\limits_{\delta_t, \beta_{it}} \left|\frac{\partial s_{it\mathbf{b}}(\delta_t;\beta_{it})}{\partial \delta_{t\mathbf{b}}}\right| = \sup\limits_{\delta_t, \beta_{it}} |s_{it\mathbf{b}}(1 - s_{it\mathbf{b}})| \leq \frac{1}{4} < e = A_1$. 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,

$$
\begin{aligned}
\frac{a_l}{l!} &\leq \frac{1}{l!} + \sum_{k=0}^{l-1} \frac{a_k}{k!} \frac{1}{(l-k)!} \\
&\leq \frac{1}{l!} + \sum_{k=0}^{l-1} A_k \frac{1}{(l-k)!} \\
&\leq \frac{1}{l!} + A_{l-1} \sum_{k=0}^{l-1} \frac{1}{(l-k)!} \\
&\leq \frac{1}{l!} + A_{l-1}(e-1) \\
&= A_l.
\end{aligned}
\tag{42}
$$

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

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

Lemma 5. *Suppose $C_{t1} \geq 2$. For any $\mathbf{b} \in \mathbf{C}_{t1}$ and $l \geq 0$,*

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

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

Proof. We prove the result by induction. For $l = 0$, the result holds trivially. For $l = 1$, the result follows directly from Lemma 4 with $l = 1$. For $l = 2$ and $l_{\mathbf{b}'} = 2$, according to Lemma 4, we have $\left| \frac{\partial^2 s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}'}^2} \right| \leq A_2 2!$. For $l = 2$ and $l_{\mathbf{b}'} = l_{\mathbf{b}''} = 1, \mathbf{b}' \neq \mathbf{b}''$:

$$\begin{aligned}
e^{\delta_{t\mathbf{b}} + \mu_{it\mathbf{b}}(\beta_{it})} &= s_{it\mathbf{b}} \sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'} + \mu_{it\mathbf{b}'}(\beta_{it})}, \\
0 &= \frac{\partial^2 s_{it\mathbf{b}''}}{\partial \delta_{t\mathbf{b}'} \partial \delta_{t\mathbf{b}''}} \sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'} + \mu_{it\mathbf{b}'}(\beta_{it})} + e^{\delta_{t\mathbf{b}'} + \mu_{it\mathbf{b}'}(\beta_{it})} \frac{\partial s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}''}} + e^{\delta_{t\mathbf{b}''} + \mu_{it\mathbf{b}''}(\beta_{it})} \frac{\partial s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}'}}, \\
\frac{\partial^2 s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}'} \partial \delta_{t\mathbf{b}''}} &= -s_{it\mathbf{b}'} \frac{\partial s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}''}} - s_{it\mathbf{b}''} \frac{\partial s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}'}},
\end{aligned} \tag{43}$$

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

Suppose that for $k = 0, ..., l-1$ the inequality holds for any $\sum_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'} = k$. First, remember that $A_l = (e-1)^l \sum_{k=0}^l \frac{1}{(e-1)^k k!}$, as defined in Lemma 4, 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 exist $\mathbf{b}'$ 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 (41) with respect to $\delta_{t\mathbf{b}}$, we obtain:

$$
\begin{aligned}
e^{\delta_{t\mathbf{b}}+\mu_{it\mathbf{b}}(\beta_{it})} &= \frac{\partial^{l_\mathbf{b}} e^{\delta_{t\mathbf{b}}+\mu_{it\mathbf{b}}(\beta_{it})}}{\partial \delta_{t\mathbf{b}}^{l_\mathbf{b}}} \\
&= \frac{\partial^{l_\mathbf{b}}\left(s_{it\mathbf{b}} \sum_{\mathbf{b}'\in\mathbf{C}_t} e^{\delta_{t\mathbf{b}'}+\mu_{it\mathbf{b}'}(\beta_{it})}\right)}{\partial \delta_{t\mathbf{b}}^{l_\mathbf{b}}} \\
&= \sum_{k=0}^{l_\mathbf{b}} C_{l_\mathbf{b}}^k \frac{\partial^k s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^k} \frac{\partial^{l_\mathbf{b}-k} \sum_{\mathbf{b}'\in\mathbf{C}_t} e^{\delta_{t\mathbf{b}'}+\mu_{it\mathbf{b}'}(\beta_{it})}}{\partial \delta_{t\mathbf{b}}^{l_\mathbf{b}-k}} \\
&= \frac{\partial^{l_\mathbf{b}} s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^{l_\mathbf{b}}} \sum_{\mathbf{b}'\in\mathbf{C}_t} e^{\delta_{t\mathbf{b}'}+\mu_{it\mathbf{b}'}(\beta_{it})} + e^{\delta_{t\mathbf{b}}+\mu_{it\mathbf{b}}(\beta_{it})} \sum_{k=0}^{l_\mathbf{b}-1} C_{l_\mathbf{b}}^k \frac{\partial^k s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^k}.
\end{aligned} \tag{44}
$$

Note that, by taking derivatives of both sides of equation (44) with respect to $\delta_{t\mathbf{b}'}$, $\mathbf{b}' \neq \mathbf{b}$, the left hand-side vanishes and we obtain:

$$
0 = \frac{\partial^{l_\mathbf{b}+l_{\mathbf{b}'}} s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^{l_\mathbf{b}} \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}} \sum_{\mathbf{b}''\in\mathbf{C}_t} e^{\delta_{t\mathbf{b}''}+\mu_{it\mathbf{b}''}(\beta_{it})} + e^{\delta_{t\mathbf{b}'}+\mu_{it\mathbf{b}'}(\beta_{it})} \sum_{k=0}^{l_{\mathbf{b}'}-1} C_{l'_\mathbf{b}}^k \frac{\partial^{l_\mathbf{b}+k} s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^{l_\mathbf{b}} \partial \delta_{t\mathbf{b}'}^k} + e^{\delta_{t\mathbf{b}}+\mu_{it\mathbf{b}}(\beta_{it})} \sum_{k=0}^{l_\mathbf{b}-1} C_{l_\mathbf{b}}^k \frac{\partial^{k+l_{\mathbf{b}'}} s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^k \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}}. \tag{45}
$$

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_{it\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_{it\mathbf{b}''}(\beta_{it})} + \sum_{\mathbf{b}'\in\mathbf{C}_{t1}} e^{\delta_{t\mathbf{b}'}+\mu_{it\mathbf{b}'}(\beta_{it})} \sum_{k=0}^{l_{\mathbf{b}'}-1} C_{l_{\mathbf{b}'}}^k \frac{\partial^{l-l_{\mathbf{b}'}+k} s_{it\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_{it\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_{it\mathbf{b}'} \sum_{k=0}^{l_{\mathbf{b}'}-1} C_{l_{\mathbf{b}'}}^k \frac{\partial^{l-l_{\mathbf{b}'}+k} s_{it\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_{it\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_{it\mathbf{b}'} \sum_{k=0}^{l_{\mathbf{b}'}-1} \frac{1}{(l_{\mathbf{b}'}-k)!} \frac{\frac{\partial^{l-l_{\mathbf{b}'}+k} s_{it\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_{it}} \left| \frac{\frac{\partial^l s_{it\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_{it}} \left| \frac{\frac{\partial^{l-l_{\mathbf{b}'}+k} s_{it\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}'}!} \right|.
\end{aligned} \tag{46}
$$

Then, applying the conclusion for any $k = 0, ..., l-1$ on the last equation in (46), we obtain:

$$\begin{aligned}
\sup_{\delta_t,\beta_{it}} \left| \frac{\frac{\partial^l s_{it\mathbf{b}}}{\prod_{\mathbf{b}'\in\mathbf{C}_{t1}} \partial\delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}}}{\prod_{\mathbf{b}'\in\mathbf{C}_{t1}} l_{\mathbf{b}'}!} \right| &\leq \sum_{\mathbf{b}'\in\mathbf{C}_{t1}} \sum_{k=0}^{l_{\mathbf{b}'}-1} \frac{1}{(l_{\mathbf{b}'}-k)!} [C_{t1}(e-1)]^{l-l_{\mathbf{b}'}+k} \\
&= [C_{t1}(e-1)]^l \sum_{\mathbf{b}'\in\mathbf{C}_{t1}} \sum_{k=1}^{l_{\mathbf{b}'}} \frac{1}{k!} [C_{t1}(e-1)]^{-k} \\
&\leq [C_{t1}(e-1)]^l C_{t1}(e^{[C_{t1}(e-1)]^{-1}} - 1) \\
&\leq [C_{t1}(e-1)]^l [e^{(e-1)^{-1}} - 1] \\
&< [C_{t1}(e-1)]^l.
\end{aligned} \tag{47}$$

Hence, the conclusion holds for $\sum_{\mathbf{b}'\in\mathbf{C}_{t1}} l_{\mathbf{b}'} = l$, and $\sup_{\delta_t,\beta_{it}} \left| \frac{\partial^l s_{it\mathbf{b}}}{\prod_{\mathbf{b}'\in\mathbf{C}_{t1}} \partial\delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}} \right| \leq [C_{t1}(e-1)]^l \prod_{\mathbf{b}'\in\mathbf{C}_{t1}} l_{\mathbf{b}'}!$ for any $l > 0$ and $\sum_{\mathbf{b}'\in\mathbf{C}_{t1}} l_{\mathbf{b}'} = l$. 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:

$$\begin{aligned}
\left| \frac{\partial^l s_{t\mathbf{b}}(\delta_t; F)}{\prod_{\mathbf{b}'\in\mathbf{C}_{t1}} \partial\delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}} \right| &\leq \int \left| \frac{\partial^l s_{it\mathbf{b}}(\delta_t; \beta_{it})}{\prod_{\mathbf{b}'\in\mathbf{C}_{t1}} \partial\delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}} \right| dF(\beta_{it}) \\
&\leq [C_{t1}(e-1)]^l \prod_{\mathbf{b}'\in\mathbf{C}_{t1}} l_{\mathbf{b}'}!
\end{aligned} \tag{48}$$

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

$$\begin{aligned}
\left| \sum_{L=0}^{\infty} \frac{1}{L!} \left[\sum_{\mathbf{b}'\in\mathbf{C}_{t1}} (\delta'_{t\mathbf{b}'} - \delta_{t\mathbf{b}'}) \frac{\partial}{\partial\delta_{t\mathbf{b}'}} \right]^L s_{t\mathbf{b}}(\delta_t; F) \right| &\leq \sum_{L=0}^{\infty} \frac{1}{L!} d^L \sum_{\sum l_{\mathbf{b}'}=L} \frac{L!}{\prod_{\mathbf{b}'\in\mathbf{C}_{t1}} l_{\mathbf{b}'}!} \left| \frac{\partial^L s_{t\mathbf{b}}(\delta_t; F)}{\prod_{\mathbf{b}'\in\mathbf{C}_{t1}} \partial\delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}} \right| \\
&\leq \sum_{L=0}^{\infty} d^L C_{t1}^L [C_{t1}(e-1)]^L,
\end{aligned} \tag{49}$$

where $d = |\delta'_t - \delta_t|$. Consequently, whenever $d < d^* = \frac{1}{C_{t1}^2(e-1)}$, the Taylor expansion (49) 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 $|\delta_t' - \delta_t| < \frac{d^*}{2}$:

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

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.

8.9 Proof of Theorem 3

The necessity part of the first statement is immediate. To prove sufficiency, note that when $(\delta_{t\mathbf{J}_t}^0)_{t \in \mathbf{T} \setminus \mathbf{T}_0} \in \Delta^{\mathrm{ID}} = \cap_{r=1}^{R} \Delta_r^{\mathrm{ID}}$, for any $r = 1, .., R$ there exists some market $t \in \mathbf{T} \setminus \mathbf{T}_0$ such that $M_t s_t^{-1}(s_t; \Sigma_F^r) \neq \Gamma^r + g_t(\Sigma_g^r)$ 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 $S = \{\Sigma^r : r = 0, ..., R\}$. Consequently, the set of solutions to $m(\Sigma'; \mathbf{T}) = 0$ is a subset of S. 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 statement, 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

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

where Z_t^r is the zero set of function $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0)); \Sigma_F^0); \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$. 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. $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0)); \Sigma_F^0); \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$ 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_F, because $s_t(\delta_t(\Gamma); \Sigma_F)$ is a composition of two real analytic functions, $\delta_t(\Gamma) : \mathbb{R}^{J_t} \to \mathbb{R}^{C_{t1}}$ and $s_t(\cdot; \Sigma_F) : \mathbb{R}^{C_{t1}} \to (0,1)^{C_{t1}}$ (from Theorem 2), it is itself a real analytic function from $\mathbb{R}^{J_t}$ to $(0,1)^{C_{t1}}$. Moreover, because $s_t(\cdot;\Sigma_F^r)$ is real analytic with respect to $\delta_t \in \mathbb{R}^{C_{t1}}$, the inverse market share function from Lemma 1, $s_t^{-1}(\cdot;\Sigma_F^r)$, is also real analytic with respect to $\mathcal{s}_t' \in (0,1)^{C_{t1}}$. Then, the composition of $M_t s_t^{-1}(\mathcal{s}_t';\Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$ and $\mathcal{s}_t' = s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0));\Sigma_F^0)$ 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 $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0));\Sigma_F^0);\Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$. There are two cases to be considered. When $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0));\Sigma_F^0);\Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$ is a constant different from zero, $Z_t^r = \emptyset$ and it has zero Lebesgue measure. Similarly, when $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0));\Sigma_F^0);\Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$ is not a constant, according to Mityagin (2015), Z_t^r has also zero Lebesgue measure.[37] This completes the proof.

8.10 Price-Setting Models Consistent with Assumption 4

Here we show that Assumption 4 is consistent with commonly employed pure components pricing models with any profile of demand synergies (substitutability and/or complementarity).

To simplify notation, in this Appendix we drop the market index t. 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 components $p_{\mathbf{b}} = \sum_{j\in\mathbf{b}} p_j$, where each p_j is chosen by the firm that owns it. Then, the profit function of firm f takes the following form:

$$\pi_f(p_f, p_{-f}) = \sum_{j\in f} s_{j.}(p_{\mathbf{J}})(p_j - c_j), \tag{50}$$

where $s_{j.}(p_{\mathbf{J}}) = \sum_{\mathbf{b}:\mathbf{b}\ni j} s_{\mathbf{b}}(p_{\mathbf{J}})$ is the product-level market share function of product j and $p_{\mathbf{J}} = (p_1, ..., p_J)$. 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:

$$\left[\frac{\partial s_{\mathbf{J}.}}{\partial p_{\mathbf{J}}} \odot \Omega\right](p_{\mathbf{J}} - c_{\mathbf{J}}) + s_{\mathbf{J}.}(p_{\mathbf{J}}) = 0, \tag{51}$$

where $\odot$ denotes the Hadamard product, or element-by-element multiplication, $s_{\mathbf{J}.} = (s_{j.}(p_{\mathbf{J}}))_{j\in\mathbf{J}}$ is the vector of product-level market share functions, $p_{\mathbf{J}} = (p_j)_{j\in\mathbf{J}}$, and $c_{\mathbf{J}} = (c_j)_{j\in\mathbf{J}}$. Given different configurations of the ownership matrix, (51) specialize to different market structures

[37] More generally, the zero set of a non-constant real 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 has zero Lebesgue measure. For details, see the second statement of Theorem 6.3.3 (Lojasiewicz's Structure Theorem for Varieties) from Krantz and Parks (2002).

such as monopoly, duopoly, or oligopoly.

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

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

We now show that for any ownership matrix, $\frac{\partial s_{\mathbf{J}.}}{\partial p_{\mathbf{J}}} \odot \Omega$ is invertible. Let $p = (p_{\mathbf{J}}, (p_{\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_2})$ denote the vector of prices for all single products and 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_t^1 introduced prior to Theorem 3, we can write:

$$\begin{aligned} \frac{\partial s_{\mathbf{J}.}}{\partial p_{\mathbf{J}}} &= \int [\mathbf{I} \;\; M_t^{1\mathrm{T}}] \frac{\partial s_i(\beta_i)}{\partial p_{\mathbf{J}}} dF(\beta_i) \\ &= \int \alpha_i [\mathbf{I} \;\; \text{-}M_t^{1\mathrm{T}}] \frac{\partial s_i(\beta_i)}{\partial u_i} [\mathbf{I} \;\; \text{-}M_t^{1\mathrm{T}}]^{\mathrm{T}} dF(\beta_i), \end{aligned} \tag{52}$$

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 8.2), $\frac{\partial s_i(\beta_i)}{\partial u_i}$ is positive-definite for any β_i. Moreover, $[\mathbf{I} \;\; \text{-}M_t^{1\mathrm{T}}]$ is of full row rank and therefore $[\mathbf{I} \;\; \text{-}M_t^{1\mathrm{T}}]^{\mathrm{T}}$ is of full column rank. Consequently, $[\mathbf{I} \;\; \text{-}M_t^{1\mathrm{T}}] \frac{\partial s_i(\beta_i)}{\partial u_i} [\mathbf{I} \;\; \text{-}M_t^{1\mathrm{T}}]^{\mathrm{T}}$ 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 1's and 0's. Then, $\frac{\partial s_{\mathbf{J}.}}{\partial p_{\mathbf{J}}} \odot \Omega$ is also block diagonal. Because each block is a principal sub-matrix of $\frac{\partial s_{\mathbf{J}.}}{\partial p_{\mathbf{J}}}$, these blocks are also negative-definite. Then, $\frac{\partial s_{\mathbf{J}.}}{\partial p_{\mathbf{J}}} \odot \Omega$ is negative-definite and thus invertible.

8.11 Proof of Corollary 2

Take $\Sigma = \Sigma^r$ and $\Gamma = \Gamma^r$. Because $s_t(\delta_t'(\Gamma^r + g_t(\Gamma_g^r)); p_{t\mathbf{J}_t}', \Sigma_F^r)$ is real analytic with respect to $(\delta_{t\mathbf{J}_t}', p_{t\mathbf{J}_t}')$, then the inverse market share function, $s_t^{-1}(\mathscr{s}_t'; p_{t\mathbf{J}_t}', \Sigma_F^r)$, is real analytic with respect to $(\mathscr{s}_t', p_{t\mathbf{J}_t}')$. Consequently, $M_t s_t^{-1}(\mathscr{s}_t'; p_{t\mathbf{J}_t}', \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$ is real analytic with respect to $(\mathscr{s}_t', p_{t\mathbf{J}_t}')$. Moreover, for $\Sigma = \Sigma^0$ and $\Gamma = \Gamma^0$, $s_t(\delta_t'(\Gamma^0 + g_t(\Gamma_g^0)); p_{t\mathbf{J}_t}', \Sigma_F^0)$ is real analytic with respect to $(\delta_{t\mathbf{J}_t}', p_{t\mathbf{J}_t}')$. Then, the composition $M_t s_t^{-1}(s_t(\delta_t'(\Gamma^0 + g_t(\Sigma_g^0)); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$ is real analytic with respect to $(\delta_{t\mathbf{J}_t}', p_{t\mathbf{J}_t}')$.

We now prove Corollary 2 by contradiction. Suppose that Assumption 5 does not hold. Then, for some $r = 1, ..., R$ and $t \in \mathbf{T} \setminus \mathbf{T}_0$, there exists a set $\tilde{\mathbf{P}}_t \subset \mathbf{P}_t$ such that $\tilde{\mathbf{P}}_t$ has positive Lebesgue measure and

$$\Gamma^r + g_t(\Sigma_g^r) = M_t s_t^{-1}(s_t(\delta_t'(\Gamma^0 + g_t(\Sigma_g^0)); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r)$$

for any $\xi'_{t\mathbf{J}_t} \in \mathbb{R}^{J_t}$. We then obtain that the zero set of the real analytic function $M_t s_t^{-1}(s_t(\delta'_t(\Gamma^0 + g_t(\Sigma_g^0)); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$ is at least $\tilde{\mathbf{P}}_t \times \mathbb{R}^{J_t}$. Because the Lebesgue measure of $\tilde{\mathbf{P}}_t$ is positive, then the Lebesgue measure of $\tilde{\mathbf{P}}_t \times \mathbb{R}^{J_t}$ is also positive. According to Mityagin (2015), $M_t s_t^{-1}(s_t(\delta'_t(\Gamma^0 + g_t(\Sigma_g^0)); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$ is then constant and equal to zero on $\mathbf{P}_t \times \mathbb{R}^{J_t}$. This contradicts Ξ_r^{ID} being non-empty. This completes the proof.

8.12 Proof of Theorem 4

Here we rely on the same notation for M_t as introduced prior to Theorem 3. M_t is a matrix of dimension $C_{t2} \times C_{t1}$. Remember that C_{t2} is the number of bundles and C_{t1} the number of inside options (bundles plus single products). 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 bundles. Each row of M_i^1 identifies with -1's the product composition of the corresponding bundle. M_t^2 is instead an identity matrix $\mathbf{I}$ of dimension $C_{t2} \times C_{t2}$, with the rows corresponding to bundles.

The proof of the first statement is similar to that of Theorem 3. On the one hand, when $(\xi^0_{t\mathbf{J}_t}, c^0_{t\mathbf{J}_t})_{t\in\mathbf{T}\setminus\mathbf{T}_0} \in \Xi^{\mathrm{ID}} = \cap_{r=1}^R \Xi_r^{\mathrm{ID}}$, for any $r = 1, .., R$ there exists some market $t \in \mathbf{T} \setminus \mathbf{T}_0$ such that $M_t s_t^{-1}(\jmath_t; p_{t\mathbf{J}_t}, \Sigma_F^r) \neq \Gamma^r + g_t(\Sigma_g^r)$ 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 $S = \{\Sigma^r : r = 0, ..., R\}$. Consequently, the set of solutions to $m(\Sigma'; \mathbf{T}) = 0$ is a subset of S. 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 statement, 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

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

where Z_t^{+r} is the zero set of $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$ such that $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0)); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r) = 0$ for some $p_{t\mathbf{J}_t} \in p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$. It then suffices to show that there exists a $t \in \mathbf{T} \setminus \mathbf{T}_0$ such that the Lebesgue measure of Z_t^{+r} is zero.

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

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

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 Theorem 1.16-2 of Ciarlet (2013) and Fubini's Theorem, we obtain:

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

Suppose that Assumption 5 holds. Denote by $\Delta_{t\mathbf{J}_t} = (\Delta_{tj}(x_{tj}, p_{tj}))_{j \in \mathbf{J}_t}$. Because $\delta_{t\mathbf{J}_t}(\Delta_{t\mathbf{J}_t}, \xi_{t\mathbf{J}_t}) = \Delta_{t\mathbf{J}_t} + \xi_{t\mathbf{J}_t}$, given $p_{t\mathbf{J}_t}$ (and therefore $\Delta_{t\mathbf{J}_t}$) and by applying Theorem 2, we obtain that the market share function $s_t(\delta_t(\Gamma' + g_t(\Sigma'_g)); p_{t\mathbf{J}_t}, \Sigma'_F)$ is also real analytic with respect to $\xi_{t\mathbf{J}_t} \in \mathbb{R}^{J_t}$. Then, given $p_{t\mathbf{J}_t}$, by the Inverse Function Theorem for real analytic functions, $s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0)); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r)$ is real analytic with respect to $\xi_{t\mathbf{J}_t}$, and therefore $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0)); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$ is real analytic with respect to $\xi_{t\mathbf{J}_t}$. For each $r = 1, ..., R$, we focus on the market $t \in \mathbf{T} \setminus \mathbf{T}_0$ that satisfies Assumption 5: for any $p_{t\mathbf{J}_t} \in \mathbf{P}_t$, there exists $\xi_{t\mathbf{J}_t} \in D_{t\xi}$ such that $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0)); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r) \neq \Gamma^r + g_t(\Sigma_g^r)$, i.e., $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0)); p_{t\mathbf{J}_t}, \Sigma_F^0); \ p_{t\mathbf{J}_t}, \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$ is not always equal to zero in $D_{t\xi}$. Similar to the proof of the second statement of Theorem 3, $\{\xi_{t\mathbf{J}_t} : \ M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0)); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r) = \Gamma^r + g_t(\Sigma_g^r)\}$ has thus zero Lebesgue measure in $D_{t\xi}$ and

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

It then follows that

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

and finally $me(Z_t^{+r}) \le 0$. Consequently, $me(Z_t^{+r}) = 0$. This completes the proof.

8.13 Proof of Theorem 5

We first introduce some notation. Denote the collection of demand synergies that can rationalize the observed product-level market shares $\mathscr{s}_{t\mathbf{J}_t\cdot}$ in market t by $\bar{\Theta}_\Gamma^t(\Sigma') = \{\Gamma' : \exists \delta'_{t\mathbf{J}_t} \in \mathbb{R}^{J_t}$ such that $s_{t\mathbf{J}_t\cdot}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma') = \mathscr{s}_{t\mathbf{J}_t\cdot}\}$ and across all T markets by $\bar{\Theta}_\Gamma = \cup_{\Sigma' \in \Theta_\Sigma} \cap_{t=1}^T \bar{\Theta}_\Gamma^t(\Sigma')$. Define also $\bar{\Theta}_\Sigma = \{\Sigma' : \cap_{t=1}^T \bar{\Theta}_\Gamma^t(\Sigma') \neq \emptyset\}$. $\bar{\Theta}_\Gamma^t$ and $\bar{\Theta}_\Sigma$ collect the values of Γ' and of Σ' that can rationalize the observed product-level market shares. When $\Sigma' \notin \bar{\Theta}_\Sigma$ or $\Gamma' \notin \cap_{t=1}^T \bar{\Theta}_\Gamma^t(\Sigma')$, then there exists no $\delta'_{t\mathbf{J}_t}$ such that $s_{t\mathbf{J}_t\cdot}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma') = \mathscr{s}_{t\mathbf{J}_t\cdot}$ for any $t = 1, ..., T$ (i.e., the demand inverse is not defined at (Γ', Σ')). The remainder of the proof focuses on the case of $\Sigma' \in \bar{\Theta}_\Sigma$ and $\Gamma' \in \cap_{t=1}^T \bar{\Theta}_\Gamma^t(\Sigma')$ (i.e., the demand inverse is defined at (Γ', Σ')), and in particular on showing the uniqueness of the corresponding $\delta'_{t\mathbf{J}_t}$.

We rely on the same notation for M_t as introduced prior to Theorem 3 and 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 5 for market t.

Denote by $\mathbf{1}$ a vector of 1's and 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 bundles consistent with the observed product-level market shares, $\mathscr{s}_{t\mathbf{J}_t\cdot}$. Given any $\mathscr{s}'_{t\mathbf{C}_{t2}} \in \mathbf{S}_{t2}(\mathscr{s}_{t\mathbf{J}_t\cdot})$ and observed product-level market shares $\mathscr{s}_{t\mathbf{J}_t\cdot}$, we can construct an admissible vector of market shares $\mathscr{s}'_t = ((\mathscr{s}'_{tj})_{j\in\mathbf{J}_t}, \mathscr{s}'_{t\mathbf{C}_{t2}})$, 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 $\delta'_t \in \mathbb{R}^{C_{t1}}$:

$$\begin{aligned}
\delta'_t &= ((\delta'_{tj})_{j\in\mathbf{J}_t}, (\delta'_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_{t2}})^{\mathrm{T}} \\
&= s_t^{-1}(\mathscr{s}'_t; \Sigma'_F) \\
&= s_t^{-1}((\mathscr{s}_{tj\cdot} - \sum_{\mathbf{b}\in\mathbf{C}_{t2}:j\in\mathbf{b}} \mathscr{s}'_{t\mathbf{b}})_{j\in\mathbf{J}_t}, \mathscr{s}'_{t\mathbf{C}_{t2}}; \Sigma'_F), \text{ where} \\
\delta'_{tj} &= s_{tj}^{-1}((\mathscr{s}_{tj\cdot} - \sum_{\mathbf{b}\in\mathbf{C}_{t2}:j\in\mathbf{b}} \mathscr{s}'_{t\mathbf{b}})_{j\in\mathbf{J}_t}, \mathscr{s}'_{t\mathbf{C}_{t2}}; \Sigma'_F), \\
\delta'_{t\mathbf{b}} &= s_{t\mathbf{b}}^{-1}((\mathscr{s}_{tj\cdot} - \sum_{\mathbf{b}\in\mathbf{C}_{t2}:j\in\mathbf{b}} \mathscr{s}'_{t\mathbf{b}})_{j\in\mathbf{J}_t}, \mathscr{s}'_{t\mathbf{C}_{t2}}; \Sigma'_F).
\end{aligned} \tag{53}$$

Using the matrix M_t, we can recover an admissible Γ'_t from δ'_t by:

$$\Gamma'_t + g_t(\Sigma'_g) = M_t\delta'_t,$$

$$\begin{aligned}
\Gamma'_t &= M_t\delta'_t - g_t(\Sigma'_g) \\
&= M_t s_t^{-1}((\mathscr{s}_{tj\cdot} - \sum_{\mathbf{b}\in\mathbf{C}_{t2}:j\in\mathbf{b}} \mathscr{s}'_{t\mathbf{b}})_{j\in\mathbf{J}_t}, \mathscr{s}'_{t\mathbf{C}_{t2}}; \Sigma'_F) - g_t(\Sigma'_g) \\
&= M_t s_t^{-1}(M_t^{\mathrm{T}}\mathscr{s}'_{t\mathbf{C}_{t2}} + (\mathscr{s}_{t\mathbf{J}_t\cdot}^{\mathrm{T}}, 0, ..., 0)^{\mathrm{T}}; \Sigma'_F) - g_t(\Sigma'_g).
\end{aligned} \tag{54}$$

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 (53) holds. We now compute from (54) the derivative of $\Gamma'_t = \Gamma_t(\mathscr{s}'_{t\mathbf{C}_{t2}}; \mathscr{s}_{t\mathbf{J}_t \cdot}, \Sigma')$ with respect to $\mathscr{s}'_{t\mathbf{C}_{t2}}$:

$$\begin{aligned} \frac{d\Gamma_t}{d\mathscr{s}'_{t\mathbf{C}_{t2}}} &= M_t \frac{\partial s_t^{-1}}{\partial \mathscr{s}'_t}(M_t^{\mathrm{T}} \mathscr{s}'_{t\mathbf{C}_{t2}} + (\mathscr{s}^{\mathrm{T}}_{t\mathbf{J}_t \cdot}, 0, ..., 0)^{\mathrm{T}}; \Sigma'_F) M_t^{\mathrm{T}} \\ &= M_t \left[\frac{\partial s_t}{\partial \delta'_t}(\delta'_t; \Sigma'_F) \right]^{-1} M_t^{\mathrm{T}}. \end{aligned} \tag{55}$$

Because $\frac{\partial s_t}{\partial \delta'_t}(\delta'_t; \Sigma'_F)$ 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 $\mathscr{s}'_{t\mathbf{C}_{t2}} \in \mathbf{S}_{t2}(\mathscr{s}_{t\mathbf{J}_t \cdot})$.[38] Note that $\mathbf{S}_{t2}(\mathscr{s}_{t\mathbf{J}_t \cdot})$ is convex. According to Theorem 6 by Gale and Nikaido (1965), p. 88, $\Gamma'_t = \Gamma_t(\mathscr{s}'_{t\mathbf{C}_{t2}}; \mathscr{s}_{t\mathbf{J}_t \cdot}, \Sigma')$ 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 $\Gamma'_t \in \bar{\Theta}^t_\Gamma$, 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, by plugging $\mathscr{s}'_{t\mathbf{C}_{t2}} = \tilde{s}_{t\mathbf{C}_{t2}}(\Gamma'; \mathscr{s}_{t\mathbf{J}_t \cdot}, \Sigma')$ into (53), 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 product-level market shares $\mathscr{s}_{t\mathbf{J}_t \cdot}$:

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

and determine the remaining $\delta'_{t\mathbf{b}}$ for each $\mathbf{b} \in \mathbf{C}_{t2}$ by $\delta'_{t\mathbf{b}} = \sum_{j \in \mathbf{b}} \delta'_{tj} + \Gamma'_{\mathbf{b}} + g_{t\mathbf{b}}(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma'_g)$, so that $s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F) = \mathscr{s}'_{t\mathbf{b}}$ 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' is given, according to Lemma 1, $s_t(\delta''_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F) \neq s_t(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F)$. Moreover, because also $\mathscr{s}_{t\mathbf{J}_t \cdot}$ is given, then there must exist some $\mathbf{b} \in \mathbf{C}_{t2}$ for which $s_{t\mathbf{b}}(\delta''_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F) \neq s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F)$. This contradicts $\tilde{s}_{t\mathbf{C}_{t2}}(\Gamma'; \mathscr{s}_{t\mathbf{J}_t \cdot}, \Sigma')$ being a function of Γ'.

8.14 Proof of Theorem 6

We start by proving a useful Lemma. Denote the log-likelihood function evaluated at the market shares observed without sampling error by:

[38] A square matrix B is positive quasi-definite if $\frac{1}{2}(B + B^{\mathrm{T}})$ is positive-definite.

$$\ell(\delta'_{1\mathbf{J}_1}, ..., \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma') = \sum_{t=1}^{T} \sum_{\mathbf{b}\in\mathbf{C}_t} \mathscr{s}_{t\mathbf{b}} \log s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F). \tag{56}$$

Lemma 6. *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. We first show that $\ell(\delta'_{1\mathbf{J}_1}, ..., \delta'_{I\mathbf{J}_T}, \Gamma', \Sigma')$ is maximized at the true $(\delta_{1\mathbf{J}_1}, ..., \delta_{I\mathbf{J}_T}, \Gamma, \Sigma)$. Note that for any $t = 1, ..., T$ and $\mathbf{b} \in \mathbf{C}_t$, $\mathscr{s}_{t\mathbf{b}} = s_{t\mathbf{b}}(\delta_t(\Gamma + g_t(\Sigma_g)); \Sigma_F)$. Then, by using Jensen's inequality, for any $(\delta'_{1\mathbf{J}_1}, ..., \delta'_{I\mathbf{J}_T}, \Gamma', \Sigma')$ we have:

$$\begin{aligned} \ell(\delta'_{1\mathbf{J}_1}, ..., \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma') - \ell(\delta_{1\mathbf{J}_1}, ..., \delta_{T\mathbf{J}_T}, \Gamma, \Sigma) &= \sum_{t=1}^{T} \sum_{\mathbf{b}\in\mathbf{C}_t} \mathscr{s}_{t\mathbf{b}} \log \frac{s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F)}{s_{t\mathbf{b}}(\delta_t(\Gamma + g_t(\Sigma_g)); \Sigma_F)} \\ &\leq \sum_{t=1}^{T} \log \sum_{\mathbf{b}\in\mathbf{C}_t} \mathscr{s}_{t\mathbf{b}} \frac{s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F)}{s_{t\mathbf{b}}(\delta_t(\Gamma + g_t(\Sigma_g)); \Sigma_F)} \\ &\leq 0. \end{aligned} \tag{57}$$

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 $\ell(\delta'_{1\mathbf{J}_1}, ..., \delta'_{I\mathbf{J}_T}, \Gamma', \Sigma')$. According to Jensen's inequality (57), this is equivalent to having $s_{t\mathbf{b}}(\tilde{\delta}_t(\tilde{\Gamma} + g_t(\tilde{\Sigma}_g)); \tilde{\Sigma}_F) = \mathscr{s}_{t\mathbf{b}}$ 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 3 rules out the possibility of having any other $\tilde{\Sigma}$ different from Σ for which system (9) holds, giving rise to a contradiction. □

We assume the following regularity conditions.

1. θ is an interior point of Θ;
2. $g_t(\Sigma'_g)$ is twice continuously differentiable with respect to Σ'_g, and the market share function $s_{t\mathbf{b}}(\delta'_t; \Sigma'_F)$, $t = 1, ..., T$ and $\mathbf{b} \in \mathbf{C}_t$, is twice continuously differentiable with respect to (δ'_t, Σ'_F);
3. $\sqrt{I}(\hat{\mathscr{s}}_t - \mathscr{s}_t) \xrightarrow{d} \mathcal{N}(0, \Omega_t)$ independently for $t = 1, ..., T$, where Ω_t is positive-definite;
4. $\sum_{t=1}^{T} G_t \Omega_t G_t^{\mathrm{T}}$ 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.

Condition 3 is compatible with cases in which the individuals in market t make independent purchase decisions. Condition 4 can be obtained when G_t is a full row rank matrix for each $t = 1, ..., T$. Define $\ell^c(\Gamma', \Sigma')$ on the basis of (56):

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

Throughout the proof, we assume that $\delta_{t\mathbf{J}_t\cdot}(\mathscr{s}_{t\mathbf{J}_t\cdot}; \Gamma', \Sigma')$ and $\delta_{t\mathbf{J}_t\cdot}(\hat{\mathscr{s}}_{t\mathbf{J}_t\cdot}; \Gamma', \Sigma')$ exist. As discussed in the main text, existence can always be verified numerically during estimation. Provided existence, then Theorem 5 guarantees that $\delta_{t\mathbf{J}_t\cdot}(\cdot; \Gamma', \Sigma')$ is a global bijection. 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. (Γ, Σ) is the unique maximizer of $\ell^c(\Gamma', \Sigma')$ 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 6 guarantees that 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. Theorem 5 then implies that (Γ, Σ) is the unique maximizer of $\ell^c(\Gamma', \Sigma')$ in $\Theta_\Gamma \times \Theta_\Sigma$.

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

3. $\ell^c(\Gamma', \Sigma')$ is continuous with respect to (Γ', Σ') in $\Theta_\Gamma \times \Theta_\Sigma$. According to regularity condition 2, for any $t = 1, ..., T$ and $\mathbf{b} \in \mathbf{C}_t$, the market share function $s_{t\mathbf{b}}(\delta'_t; \Sigma'_F)$ is twice continuously differentiable with respect to (δ'_t, Σ'_F). Remember that $\theta = (\delta_{1\mathbf{J}_1}, ..., \delta_{I\mathbf{J}_T}, \Gamma, \Sigma)$. Then, $\ell(\theta')$ in (56) is twice continuously differentiable in Θ. Moreover, the inverse market share function, $s_t^{-1}(\mathscr{s}_t; \Sigma'_F)$ is continuous with respect to $(\mathscr{s}_t, \Sigma'_F)$, and therefore continuous with respect to $((\mathscr{s}_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_{t2}}, \Sigma'_F)$. In addition, $g_t(\Sigma'_g)$ is continuously differentiable with respect to Σ'_g. Then, $\Gamma_t((\mathscr{s}_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_{t2}}, \Sigma')$, as defined in the proof of Theorem 5, is continuous with respect to $((\mathscr{s}_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_{t2}}, \Sigma')$. By applying the invertibility result from Theorem 5 and the continuous dependence with respect to Σ', we obtain that $\delta_{t\mathbf{J}_t\cdot}(\mathscr{s}_{t\mathbf{J}_t\cdot}; \Gamma', \Sigma')$ is continuous with respect to (Γ', Σ'). Combining this with the continuity of $\ell(\theta')$ in (56), we obtain the desired condition.

4. $\sup_{(\Gamma',\Sigma')\in\Theta_\Gamma\times\Theta_\Sigma} |\ell_I^c(\Gamma', \Sigma'; \hat{\mathscr{s}}_1, ..., \hat{\mathscr{s}}_T) - \ell^c(\Gamma', \Sigma')| \xrightarrow{p} 0$. Note that

$$\begin{aligned}\sup_{(\Gamma',\Sigma')\in\Theta_\Gamma\times\Theta_\Sigma} |\ell_I^c(\Gamma', \Sigma'; \hat{\mathscr{s}}_1, ..., \hat{\mathscr{s}}_T) - \ell^c(\Gamma', \Sigma')| &\le \sup_{\theta'\in\Theta} |\ell_I(\theta'; \hat{\mathscr{s}}_1, ..., \hat{\mathscr{s}}_T) - \ell(\theta')| \\ &+ \sup_{(\Gamma',\Sigma')\in\Theta_\Gamma\times\Theta_\Sigma} |\ell((\delta_{t\mathbf{J}_t\cdot}(\hat{\mathscr{s}}_{t\mathbf{J}_t\cdot}; \Gamma', \Sigma'))_{t=1}^T, \Gamma', \Sigma') - \ell((\delta_{t\mathbf{J}_t\cdot}(\mathscr{s}_{t\mathbf{J}_t\cdot}; \Gamma', \Sigma'))_{t=1}^T, \Gamma', \Sigma')|\end{aligned}$$

First, we prove that $\sup_{\theta' \in \Theta} |\ell_I(\theta'; \hat{\jmath}_1, ..., \hat{\jmath}_T) - \ell(\theta')| \xrightarrow{p} 0$. To see this, note that:

$$
\begin{aligned}
&\sup_{\theta' \in \Theta} |\ell_I(\theta'; \hat{\jmath}_1, ..., \hat{\jmath}_T) - \ell(\theta')| \\
&= \sup_{\theta' \in \Theta} \left| \sum_{t=1}^{T} \sum_{\mathbf{b} \in \mathbf{C}_t} \hat{\jmath}_{t\mathbf{b}} \log s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F) - \sum_{t=1}^{T} \sum_{\mathbf{b} \in \mathbf{C}_t} \jmath_{t\mathbf{b}} \log s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F) \right| \\
&\leq \sup_{\substack{\theta' \in \Theta \\ t=1,...,T, \mathbf{b} \in \mathbf{C}_t}} |\log s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F)| \sum_{t=1}^{T} \sum_{\mathbf{b} \in \mathbf{C}_t} |\hat{\jmath}_{t\mathbf{b}} - \jmath_{t\mathbf{b}}|.
\end{aligned}
\tag{58}
$$

Because $\log s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F)$ is continuous in Θ and Θ is compact, $\log s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F)$ is uniformly bounded in Θ. Moreover, because both the number of markets, T, and C_t are finite,

$$
\sup_{\substack{\theta' \in \Theta \\ t=1,...,T, \mathbf{b} \in \mathbf{C}_t}} |\log s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F)| < \infty.
$$

Note that $\hat{\jmath}_{t\mathbf{b}} \xrightarrow{p} \jmath_{t\mathbf{b}}$ for $t = 1, ..., T$ and $\mathbf{b} \in \mathbf{C}_t$. Then, the right-hand side of (58) converges to zero in probability. Consequently, $\sup_{\theta' \in \Theta} |\ell_I(\theta'; \hat{\jmath}_1, ..., \hat{\jmath}_T) - \ell(\theta')| \xrightarrow{p} 0$.
Second, we prove

$$
\sup_{(\Gamma', \Sigma') \in \Theta_\Gamma \times \Theta_\Sigma} |\ell((\delta_{t\mathbf{J}_t \cdot}(\hat{\jmath}_{t\mathbf{J}_t \cdot}; \Gamma', \Sigma'))_{t=1}^{T}, \Gamma', \Sigma') - \ell((\delta_{t\mathbf{J}_t \cdot}(\jmath_{t\mathbf{J}_t \cdot}; \Gamma', \Sigma'))_{t=1}^{T}, \Gamma', \Sigma')| \xrightarrow{p} 0. \tag{59}
$$

Note that for each t, $\delta_{t\mathbf{J}_t \cdot}(\jmath'_{t\mathbf{J}_t \cdot}; \Gamma', \Sigma')$ is uniformly continuous with respect to $(\jmath'_{t\mathbf{J}_t \cdot}, \Gamma', \Sigma')$ in a compact set $U_{\jmath_{t\mathbf{J}_t \cdot}} \times \Theta_\Gamma \times \Theta_\Sigma$, where $U_{\jmath_{t\mathbf{J}_t \cdot}}$ is a compact neighbourhood of $\jmath_{t\mathbf{J}_t \cdot}$. Moreover, $\ell(\theta')$ is uniformly continuous with respect to $\theta' \in \Theta$. Consequently, $\ell((\delta_{t\mathbf{J}_t \cdot}(\jmath'_{t\mathbf{J}_t \cdot}; \Gamma', \Sigma'))_{t=1}^{T}, \Gamma', \Sigma')$ is uniformly continuous with respect to $((\jmath'_{t\mathbf{J}_t \cdot})_{t=1}^{T}, \Gamma', \Sigma')$. Because $\hat{\jmath}_{t\mathbf{J}_t \cdot} \xrightarrow{p} \jmath_{t\mathbf{J}_t \cdot}$ for $t = 1, ..., T$, we obtain (59) and finally proved the desired condition.

According to Theorem 2.1 by Newey and McFadden (1994), the four conditions verified above guarantee the consistency of $(\hat{\Gamma}, \hat{\Sigma})$. By applying the invertibility result from Theorem 5 and Slutsky's Theorem, $\hat{\theta}_\delta$ is also consistent. This completes the proof of consistency. The proof of asymptotic normality is based on Theorem 3.1 by Newey and McFadden (1994), according to which we need to verify the following six conditions.

1. $(\hat{\Gamma}, \hat{\Sigma}) \xrightarrow{p} (\Gamma, \Sigma)$. 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. $\ell_I^c(\Gamma', \Sigma'; \hat{\jmath}_1, ..., \hat{\jmath}_T)$ is twice continuously differentiable in $\Theta_\Gamma \times \Theta_\Sigma$. According to regularity condition 2, the market share function $s_{t\mathbf{b}}(\delta'_t; \Sigma'_F)$, $t = 1, ..., T$ and of $\mathbf{b} \in \mathbf{C}_t$, is twice continuously differentiable with respect to (δ_t', Σ_F'), the inverse market share function $s_t^{-1}(\jmath_t'; \Sigma_F')$ is thus twice continuously differentiable with respect to $(\jmath_t'; \Sigma_F')$. Moreover, $g_t(\Sigma_g')$ is twice continuously differentiable with respect to Σ_g'. As a consequence, by applying the invertibility result from Theorem 5, we obtain that $\delta_{t\mathbf{J}_t\cdot}(\jmath_{t\mathbf{J}_t\cdot}'; \Gamma', \Sigma')$ is twice continuously differentiable with respect to $(\jmath_{t\mathbf{J}_t\cdot}', \Gamma', \Sigma')$. Because $\ell_I^c(\Gamma', \Sigma'; \hat{\jmath}_1, ..., \hat{\jmath}_T)$ is a composition of $\ell_I(\delta_{t\mathbf{J}_t}, \Gamma', \Sigma'; \hat{\jmath}_1, ..., \hat{\jmath}_T)$ and of $\delta_{t\mathbf{J}_t\cdot}(\jmath_{t\mathbf{J}_t\cdot}'; \Gamma', \Sigma')$, and both functions are twice continuously differentiable, $\ell_I^c(\Gamma', \Sigma'; \hat{\jmath}_1, ..., \hat{\jmath}_T)$ 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:

$$\begin{aligned}\sqrt{I}\frac{\partial \ell_I^c}{\partial(\Gamma', \Sigma')} &= \sqrt{I}\sum_{t=1}^{T}\frac{\partial \delta_{t\mathbf{J}_t\cdot}}{\partial(\Gamma', \Sigma')}\frac{\partial \ell_I}{\partial \delta_{t\mathbf{J}_t}} + \frac{\partial \ell_I}{\partial(\Gamma', \Sigma')} \\ &= \left[\left(\frac{\partial \delta_{t\mathbf{J}_t\cdot}}{\partial(\Gamma', \Sigma')}\right)_{t=1,...,T} \quad \mathbf{I}\right]\sqrt{I}\frac{\partial \ell_I}{\partial \theta'}\end{aligned} \tag{60}$$

It suffices to prove that $\sqrt{I}\frac{\partial \ell_I}{\partial \theta'}$ converges to a centered normal distribution at $\theta' = \theta$.

Define $\ell^t(\theta') = \sum_{\mathbf{b}\in\mathbf{C}_t} \jmath_{t\mathbf{b}} \log s_{t\mathbf{b}}(\delta_t'(\Gamma' + g_t(\Sigma_g')); \Sigma_F')$. 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,

$$\begin{aligned}\sqrt{I}\frac{\partial \ell_I}{\partial \theta'}\bigg|_{\theta'=\theta} &= \sqrt{I}\sum_{t=1}^{T}\sum_{\mathbf{b}\in\mathbf{C}_t}\hat{\jmath}_{t\mathbf{b}}\frac{\log \partial s_{t\mathbf{b}}}{\partial \theta'}\bigg|_{\theta'=\theta} \\ &= \sqrt{I}\sum_{t=1}^{T}\left[\sum_{\mathbf{b}\in\mathbf{C}_t}\hat{\jmath}_{t\mathbf{b}}\frac{\partial \log s_{t\mathbf{b}}}{\partial \theta'}\bigg|_{\theta'=\theta} - \frac{\partial \ell^t}{\partial \theta'}\bigg|_{\theta'=\theta}\right] \\ &= \sqrt{I}\sum_{t=1}^{T}\left[\sum_{\mathbf{b}\in\mathbf{C}_{t1}}[\hat{\jmath}_{t\mathbf{b}} - \jmath_{t\mathbf{b}}]\frac{\partial \log s_{t\mathbf{b}}}{\partial \theta'}\bigg|_{\theta'=\theta} + [\hat{\jmath}_{t0} - \jmath_{t0}]\frac{\partial \log s_{t0}}{\partial \theta'}\bigg|_{\theta'=\theta}\right] \\ &= \sum_{t=1}^{T}\sum_{\mathbf{b}\in\mathbf{C}_{t1}}\sqrt{I}\,[\hat{\jmath}_{t\mathbf{b}} - \jmath_{t\mathbf{b}}]\left[\frac{\partial \log s_{t\mathbf{b}}}{\partial \theta'} - \frac{\partial \log s_{t0}}{\partial \theta'}\right]\bigg|_{\theta'=\theta} \\ &= \sum_{t=1}^{T}\left(\left[\frac{\partial \log s_{t\mathbf{b}}}{\partial \theta'} - \frac{\partial \log s_{t0}}{\partial \theta'}\right]\bigg|_{\theta'=\theta}\right)_{\mathbf{b}\in\mathbf{C}_{t1}}\sqrt{I}\,[\hat{\jmath}_t - \jmath_t]\,.\end{aligned} \tag{61}$$

where $\mathbf{I}$ denotes the identity matrix. 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, we obtain that $\sqrt{I}\frac{\partial \ell_I}{\partial \theta'}\big|_{\theta'=\theta} \xrightarrow{d} \mathcal{N}(0, \sum_{t=1}^{T} G_t\Omega_t G_t^{\mathrm{T}})$, where $\sum_{t=1}^{T} G_t\Omega_t G_t^{\mathrm{T}}$ is positive-definite according to regularity condition 4. As a consequence, $\sqrt{I}\frac{\partial \ell_I^c}{\partial(\Gamma', \Sigma')}\big|_{(\Gamma', \Sigma')=(\Gamma, \Sigma)}$ converges to a centered normal distribution.

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

$$\begin{aligned} H(\Gamma',\Sigma') &= \frac{\partial^2\ell((\delta_{t\mathbf{J}_t\cdot}(\jmath_{t\mathbf{J}_t\cdot};\Gamma',\Sigma'))_{t=1}^T,\Gamma',\Sigma')}{\partial(\Gamma',\Sigma')^2} \\ &= \sum_{t=1}^T\sum_{\mathbf{b}\in\mathbf{C}_t}\jmath_{t\mathbf{b}}\frac{\partial^2\log s_{t\mathbf{b}}(\delta_t(\delta_{t\mathbf{J}_t\cdot}(\jmath_{t\mathbf{J}_t\cdot};\Gamma',\Sigma'),\Gamma'+g_t(\Sigma'_g));\Sigma'_F)}{\partial(\Gamma',\Sigma')^2}, \end{aligned} \tag{62}$$

where $\delta_t(\delta_{t\mathbf{J}_t\cdot}(\jmath_{t\mathbf{J}_t\cdot};\Gamma',\Sigma'),\Gamma'+g_t(\Sigma'_g))=(\delta_{t\mathbf{J}_t\cdot}(\jmath_{t\mathbf{J}_t\cdot};\Gamma',\Sigma'),(\sum_{j\in\mathbf{b}}\delta_{tj\cdot}(\jmath_{t\mathbf{J}_t\cdot};\Gamma',\Sigma')+\Gamma'_\mathbf{b}+g_{t\mathbf{b}}(\Sigma'_g))_{\mathbf{b}\in\mathbf{C}_{t2}})$. Under regularity condition 2, $H(\Gamma',\Sigma')$ is continuous in $\Theta_\Gamma\times\Theta_\Sigma$. Note that, similarly to (58), we have:

$$\begin{aligned} &\sup_{(\Gamma',\Sigma')\in\Theta_\Gamma\times\Theta_\Sigma}\left|\frac{\partial^2\ell_I^c}{\partial(\Gamma',\Sigma')^2}(\Gamma',\Sigma')-H(\Gamma',\Sigma')\right| \\ &\leq\sum_{t=1,\dots,T,\mathbf{b}\in\mathbf{C}_t}\sup_{(\Gamma',\Sigma')\in\Theta_\Gamma\times\Theta_\Sigma}\left|\frac{\partial^2\log s_{t\mathbf{b}}(\delta_t(\delta_{t\mathbf{J}_t\cdot}(\jmath_{t\mathbf{J}_t\cdot};\Gamma',\Sigma'),\Gamma'+g_t(\Sigma'_g));\Sigma'_F)}{\partial(\Gamma',\Sigma')^2}\right||\hat{\jmath}_{t\mathbf{b}}-\jmath_{t\mathbf{b}}| \\ &\quad+\sum_{t=1,\dots,T,\mathbf{b}\in\mathbf{C}_t}\hat{\jmath}_{t\mathbf{b}}\sup_{(\Gamma',\Sigma')\in\Theta_\Gamma\times\Theta_\Sigma}\left|\frac{\partial^2\log\left[s_{t\mathbf{b}}(\delta_t(\delta_{t\mathbf{J}_t\cdot}(\jmath_{t\mathbf{J}_t\cdot};\Gamma',\Sigma'),\Gamma'+g_t(\Sigma'_g));\Sigma'_F)-s_{t\mathbf{b}}(\delta_t(\delta_{t\mathbf{J}_t\cdot}(\hat{\jmath}_{t\mathbf{J}_t\cdot};\Gamma',\Sigma'),\Gamma'+g_t(\Sigma'_g));\Sigma'_F)\right]}{\partial(\Gamma',\Sigma')^2}\right| \end{aligned} \tag{63}$$

Due to the twice continuous differentiability of $s_{t\mathbf{b}}(\delta'_t;\Sigma')$ and of $\delta_{t\mathbf{J}_t\cdot}(\jmath_{t\mathbf{J}_t\cdot};\Gamma',\Sigma')$ in the compact set $\Theta_\Gamma\times\Theta_\Sigma$, for $t=1,...,T$ and $\mathbf{b}\in\mathbf{C}_t$, we have:

$$\sup_{(\Gamma',\Sigma')\in\Theta_\Gamma\times\Theta_\Sigma}\left|\frac{\partial^2\log s_{t\mathbf{b}}(\delta_t(\delta_{t\mathbf{J}_t\cdot}(\jmath_{t\mathbf{J}_t\cdot};\Gamma',\Sigma'),\Gamma'+g_t(\Sigma'_g));\Sigma'_F)}{\partial(\Gamma',\Sigma')^2}\right|<\infty.$$

Because $\hat{\jmath}_t\xrightarrow{p}\jmath_t$ for $t=1,...,T$, then the first part on the right-hand side of (63) converges to zero in probability. For the second part, note that $s_{t\mathbf{b}}(\delta'_t(\Gamma'+g_t(\Sigma'_g));\Sigma_F)$ is twice continuously differentiable with respect to $(\delta'_{\mathbf{J}_t},\Gamma',\Sigma')$ and that $\delta_{t\mathbf{J}_t\cdot}(\jmath'_{t\mathbf{J}_t\cdot};\Gamma',\Sigma')$ is twice continuously differentiable with respect to $(\jmath'_{t\mathbf{J}_t\cdot},\Gamma',\Sigma')$ in a compact set $U_{\jmath_{t\mathbf{J}_t\cdot}}\times\Theta_\Gamma\times\Theta_\Sigma$, where $U_{\jmath_{t\mathbf{J}_t\cdot}}$ is a compact neighbourhood of $\jmath_{t\mathbf{J}_t\cdot}$, for $t=1,...,T$ and $\mathbf{b}\in\mathbf{C}_{t1}$, we then obtain that $\frac{\partial^2\log s_{t\mathbf{b}}(\delta_t(\delta_{t\mathbf{J}_t\cdot}(\jmath_{t\mathbf{J}_t\cdot};\Gamma',\Sigma'),\Gamma'+g_t(\Sigma'_g));\Sigma'_F)}{\partial(\Gamma',\Sigma')^2}$ is uniformly continuous in $U_{\jmath_{t\mathbf{J}_t\cdot}}\times\Theta_\Gamma\times\Theta_\Sigma$. Combining this with $\hat{\jmath}_t\xrightarrow{p}\jmath_t$ for $t=1,...,T$, we obtain that the second part on the right-hand side of (63) also converges to zero in probability. Consequently, $\sup_{(\Gamma',\Sigma')\in\Theta_\Gamma\times\Theta_\Sigma}\left|\frac{\partial^2\ell_I^c}{\partial(\Gamma',\Sigma')^2}(\Gamma',\Sigma')-H(\Gamma',\Sigma')\right|\xrightarrow{p}0$.

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

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

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

$$H(\Gamma, \Sigma) = \left[\left(\frac{\partial \delta_{t\mathbf{J}_t\cdot}}{\partial(\Gamma', \Sigma')} \right)_{t=1,\dots,T} \quad \mathbf{I} \right] \left[\frac{\partial^2 \ell(\theta')}{\partial \theta'^2} \right]_{\theta'=\theta} \left[\left(\frac{\partial \delta_{t\mathbf{J}_t\cdot}}{\partial(\Gamma', \Sigma')} \right)_{t=1,\dots,T} \quad \mathbf{I} \right]^{\mathrm{T}},$$

Because $\left[\left(\frac{\partial \delta_{t\mathbf{J}_t\cdot}}{\partial(\Gamma', \Sigma')} \right)_{t=1,\dots,T} \quad \mathbf{I} \right]$ is of full row rank and $\left[\frac{\partial^2 \ell(\theta')}{\partial \theta'^2} \right]_{\theta'=\theta}$ is non-singular according to regularity condition 5, $H(\Gamma, \Sigma)$ 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 $\sqrt{I}[(\hat{\Gamma}, \hat{\Sigma}) - (\Gamma, \Sigma)] \xrightarrow{d} \mathcal{N}(0, W_2)$. By applying the invertibility result from Theorem 5, we have:

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

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

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

we obtain

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

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

8.15 Appendix Tables

Table 7: 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 the average market share across the 83256 shopping trips with *some* RTE cereal purchases. Market shares are computed over the shopping trips observed in each store-week combination.

Online Supplement to:
“Identification and Estimation of Demand for Bundles”

Alessandro Iaria and Ao Wang[1]

This Version: January 2019
First Version: March 2019

Some Intuition about Identification

In this Appendix, we illustrate that even simple versions of model (3) raise non-trivial identification issues. First, we show 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 several markets helps identification.

Suppose there are only two products in each market t, $\mathbf{J}_t = \{1, 2\}$. The indirect utility of individuals in market t by choosing $\mathbf{b} \in \{0, 1, 2, (1, 2)\}$ is:[39]

$$
\begin{aligned}
U_{it0} &= \varepsilon_{it0}, \\
U_{it1} &= \delta_{t1} + \mu_{it1} + \varepsilon_{it1}, \\
U_{it2} &= \delta_{t2} + \mu_{it2} + \varepsilon_{it2}, \\
U_{it(1,2)} &= \delta_{t1} + \delta_{t2} + \mu_{it1} + \mu_{it2} + \Gamma_t + \zeta_{it(1,2)} + \varepsilon_{it(1,2)},
\end{aligned}
\tag{64}
$$

where the individual-specific demand synergy is $\Gamma_{it(1,2)} = \Gamma_t + \zeta_{it(1,2)}$, the vector of random coefficients $\beta_{it} = (\mu_{it1}, \mu_{it2}, \zeta_{it(1,2)})$ is distributed according to $F(\beta_{it}; \Sigma_F)$, $\Sigma_F = (\sigma, r)$, 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 $\mathbf{b} \in \{0, 1, 2, (1, 2)\}$ for each market $t = 1, ..., T$. For any given observed 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) are the unique solution to the following system:

$$
\begin{aligned}
& s_t(\delta'_{t1}, \delta'_{t2}, \delta'_{t(1,2)}(\Gamma'_t); \sigma', r') = \mathscr{s}_t \\
& \text{subject to } \delta'_{t(1,2)}(\Gamma'_t) - \delta'_{t1} - \delta'_{t2} = \Gamma'_t
\end{aligned}
\tag{65}
$$

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 bundle $(1, 2)$, $\delta'_{t(1,2)}(\Gamma'_t)$. Even in this simple example, a formal discussion of identification on the basis of system (65) would require to deal with cumbersome details, and these may prevent one from seeing the main mechanism at work. We then investigate the behaviour of a linearized version of system

[1] University of Bristol and CEPR (alessandro.iaria@bristol.ac.uk) and CREST (ao.wang@ensae.fr).

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

(65) 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 (65) around the true $((\delta_t(\Gamma_t))_{t=1}^T, \sigma, r)$:[40]

$$\begin{aligned} \delta_t'(\Gamma_t') &= \delta_t(\Gamma_t) + \left.\frac{\partial s_t^{-1}}{\partial(\sigma', r')}\right|_{(\sigma',r')=(\sigma,r)} (\sigma'-\sigma, r'-r)^{\mathrm{T}} \\ &\text{subject to } \delta_{t(1,2)}'(\Gamma_t') - \delta_{t1}' - \delta_{t2}' = \Gamma_t' \end{aligned} \tag{66}$$

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 (66) by M and by plugging in the constraint, one obtains:

$$\Gamma_t' = \Gamma_t + M\left.\frac{\partial s_t^{-1}}{\partial(\sigma', r')}\right|_{(\sigma',r')=(\sigma,r)} (\sigma'-\sigma, r'-r)^{\mathrm{T}} \tag{67}$$

for $t = 1, ..., T$. System (67) has T equations in $T+2$ unknowns, Γ_t' for $t = 1, ..., T$ and (σ', r'). The system is under-determined and (66) does not have a unique solution. One way to reduce the dimensionality of (67) is to add restrictions on Γ_t. Building on Gentzkow (2007)'s insight, we consider the case of $\Gamma_t = \Gamma$ for $t = 1, ..., T$:

$$\Gamma' = \Gamma + M\left.\frac{\partial s_t^{-1}}{\partial(\sigma', r')}\right|_{(\sigma',r')=(\sigma,r)} (\sigma'-\sigma, r'-r)^{\mathrm{T}}. \tag{68}$$

Note that because Γ' and Γ in (68) are no longer market specific, system (68) 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:

$$M\left.\left[\frac{\partial s_t^{-1}}{\partial(\sigma', r')} - \frac{\partial s_1^{-1}}{\partial(\sigma', r')}\right]\right|_{(\sigma',r')=(\sigma,r)} (\sigma'-\sigma, r'-r)^{\mathrm{T}} = 0, \tag{69}$$

of $t = 2, ..., T$. If (σ', r') is a solution to system (69), then given (σ', r') one can determine the corresponding Γ' from system (68). In turn, given (σ', r') and Γ', one can obtain the remaining parameters $(\delta_{t1}', \delta_{t2}')_{t=1}^T$ from system (66). Collectively, these $(\delta_t'(\Gamma'))_{t=1}^T$ and (σ', r') constitute a solution to system (66).

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 $\mathcal{s}_a = (\mathcal{s}_{a1}, \mathcal{s}_{a2}, \mathcal{s}_{a(1,2)})$ and $\mathcal{s}_b = (\mathcal{s}_{b1}, \mathcal{s}_{b2}, \mathcal{s}_{b(1,2)})$. Because r is assumed to be known and to

[40]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 function: for any given $\mathcal{s}_t$ and (σ', r') in a neighbourhood of (σ, r), there exists a unique δ_t' such that $s_t(\delta_t'; \sigma', r') = \mathcal{s}_t$. In addition, the dependence of $\delta_t' = s_t^{-1}(\mathcal{s}_t; \sigma', r')$ on (σ', r') is continuously differentiable.

equal zero, system (69) simplifies to the equation:

$$M\left[\frac{\partial s_b^{-1}}{\partial\sigma'}-\frac{\partial s_a^{-1}}{\partial\sigma'}\right]\Bigg|_{(\sigma',r')=(\sigma,0)}(\sigma'-\sigma)=0, \tag{70}$$

where $M=(-1,-1,1)$. Note that $\sigma'=\sigma$ is the unique solution to equation (70) as long as $M\left[\frac{\partial s_b^{-1}}{\partial\sigma'}-\frac{\partial s_a^{-1}}{\partial\sigma'}\right]\Big|_{(\sigma',r')=(\sigma,0)}\neq 0$. This condition can be re-written as:

$$\left[\frac{\partial s_{a(1,2)}^{-1}}{\partial\sigma'}-\frac{\partial s_{a1}^{-1}}{\partial\sigma'}-\frac{\partial s_{a2}^{-1}}{\partial\sigma'}\right]\Bigg|_{(\sigma',r')=(\sigma,0)}\neq\left[\frac{\partial s_{b(1,2)}^{-1}}{\partial\sigma'}-\frac{\partial s_{b1}^{-1}}{\partial\sigma'}-\frac{\partial s_{b2}^{-1}}{\partial\sigma'}\right]\Bigg|_{(\sigma',r')=(\sigma,0)}, \tag{71}$$

or equivalently as:

$$\frac{\partial\Gamma(\mathscr{s}_a;\sigma',r')}{\partial\sigma'}\Bigg|_{(\sigma',r')=(\sigma,0)}\neq\frac{\partial\Gamma(\mathscr{s}_b;\sigma',r')}{\partial\sigma'}\Bigg|_{(\sigma',r')=(\sigma,0)}. \tag{72}$$

Condition (72) 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 (72) requires some variation across markets, in the sense 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 (69) simplifies to the following equation:

$$M\left[\frac{\partial s_b^{-1}}{\partial(\sigma',r')}-\frac{\partial s_a^{-1}}{\partial(\sigma',r')}\right]\Bigg|_{(\sigma',r')=(\sigma,r)}(\sigma'-\sigma,r'-r)^{\mathrm{T}}=0. \tag{73}$$

Note that $M\left[\frac{\partial s_b^{-1}}{\partial(\sigma',r')}-\frac{\partial s_a^{-1}}{\partial(\sigma',r')}\right]\Big|_{(\sigma',r')=(\sigma,r)}$ is of size 1×2 and therefore not of full column rank. It then follows that any solution to equation (73) cannot be unique: in a neighborhood of (σ,r), there exist infinitely many (σ',r')'s such that equation (73) holds.

Figure 1 provides a 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 (65) 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 (73) has multiple (σ',r') solutions, $\mathscr{s}_a$ and $\mathscr{s}_b$ do not uniquely belong to $\mathscr{S}$. As shown in the right part of the Figure in red, for any solution to equation (73), (σ',r'), also the corresponding manifold $\mathscr{S}'$

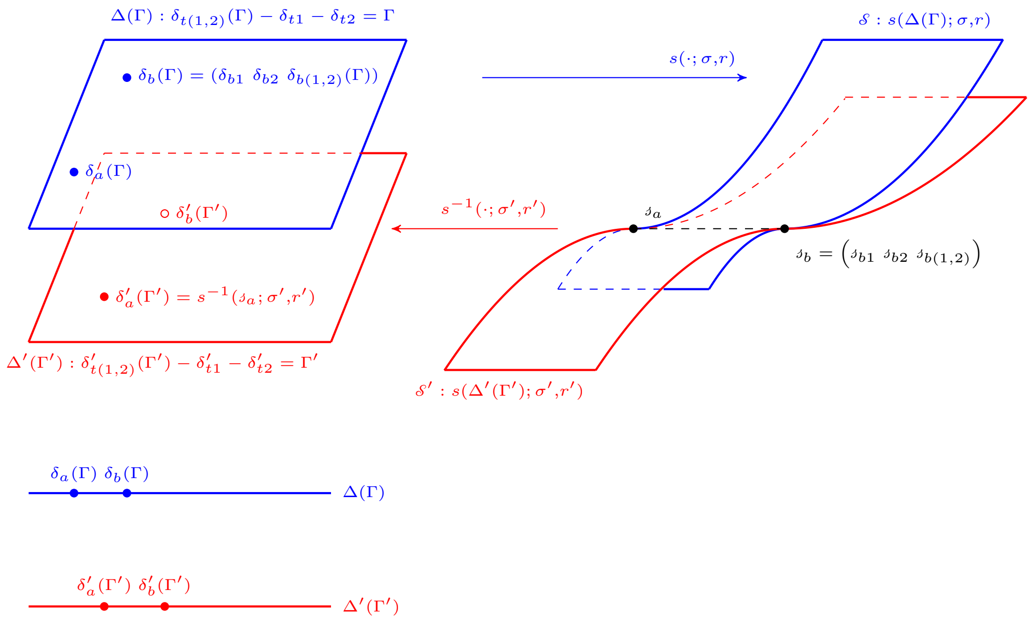

Figure 1: An example of lack of identification

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 $\delta'_b(\Gamma') \in \Delta'(\Gamma') = s^{-1}(\mathscr{S}'; \sigma', r')$ 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 (65) 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 (69) simplifies to:

$$\begin{aligned} M \left[\frac{\partial s_b^{-1}}{\partial(\sigma', r')} - \frac{\partial s_a^{-1}}{\partial(\sigma', r')} \right]_{(\sigma', r') = (\sigma, r)} (\sigma' - \sigma, r' - r)^{\mathrm{T}} = 0 \\ M \left[\frac{\partial s_c^{-1}}{\partial(\sigma', r')} - \frac{\partial s_a^{-1}}{\partial(\sigma', r')} \right]_{(\sigma', r') = (\sigma, r)} (\sigma' - \sigma, r' - r)^{\mathrm{T}} = 0. \end{aligned} \tag{74}$$

Note that (σ, r) is the unique solution to linear system (74) and the model is identified as long as the 2×2 matrix

$$\begin{bmatrix} M \left(\frac{\partial s_b^{-1}}{\partial(\sigma', r')} - \frac{\partial s_a^{-1}}{\partial(\sigma', r')} \right) \\ M \left(\frac{\partial s_c^{-1}}{\partial(\sigma', r')} - \frac{\partial s_a^{-1}}{\partial(\sigma', r')} \right) \end{bmatrix}_{(\sigma', r') = (\sigma, r)} \tag{75}$$

is of full column rank. In Example 2, the corresponding matrix in equation (73) 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

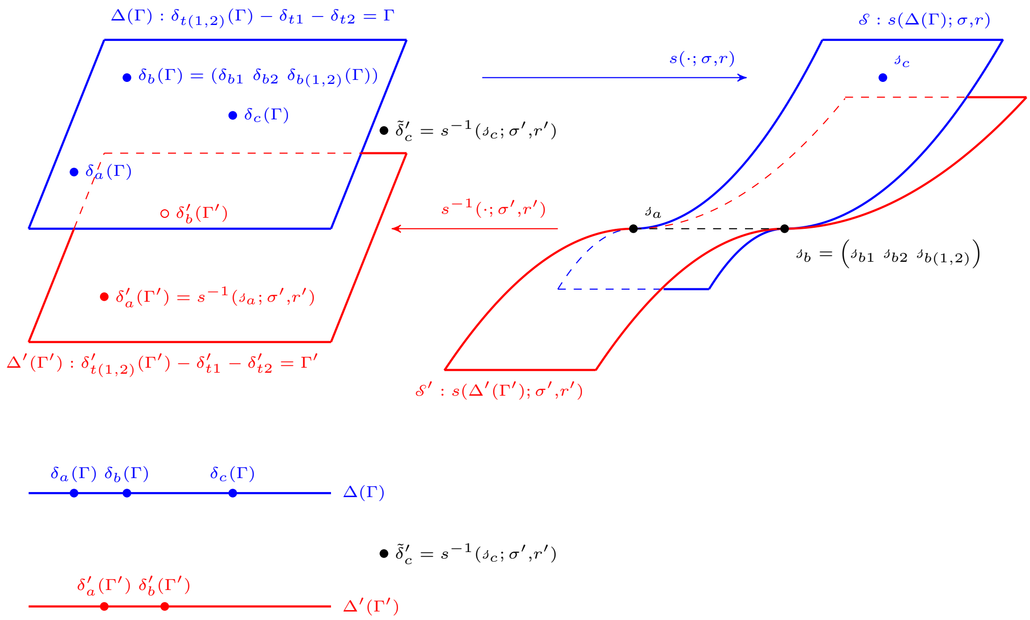

Figure 2: An example of identification

full column rank of matrix (75). The full column rank condition for the 2×2 matrix (75) generalizes identification condition (72) from Example 1.

Figure 2 provides a 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 2 is similar to that of Figure 1, with the exception of the additional $\delta_c(\Gamma) \in \Delta(\Gamma)$ and the corresponding $\mathscr{s}_c \in \mathscr{S}$. Differently from Example 2, the additional $\mathscr{s}_c$ and the full column rank of (75) guarantee that there is no manifold $\mathscr{S}'$ other than $\mathscr{S} = s(\Delta(\Gamma); \sigma, r)$ 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 2, 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 (65) and the model is identified.

These three examples highlight two general points about the identification of model (3). First, as condition (75) illustrates, the task of recovering the full set of structural parameters reduces to that of identifying the parameters of the distribution of random coefficients Σ_F. This directly follows from two features of system (65): 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 $\delta_{t(1,2)}(\Gamma) - \delta_{t1} - \delta_{t2} = \Gamma$, $t = 1, ..., T$.[41] Second, as system (69) illustrates, whenever the dimension of Σ_F does not depend on the number of markets T, adding markets to the dataset will help identification. Identification requires matrix (75) 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_F (i.e., number of columns), and identification can hardly be achieved. Differently, by adding one market to the dataset, matrix (75) 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_F only contains one parameter, σ, rather than two, σ and r.

Global Identification with Mixed Bundling Pricing

Denote by $\mathbf{B}_t \subseteq \mathbf{C}_{t2}$ the set of bundles whose single products all belong to the same firm and by B_t its cardinality. Note that, by definition, the bundles $\mathbf{b} \in \mathbf{B}_t$ can have $p_{t\mathbf{b}} \neq 0$, while $\mathbf{b}' \in \mathbf{C}_{t2}/\mathbf{B}_t$ must have $p_{t\mathbf{b}'} = 0$. Define the following vectors of prices: the prices of single products and the price surcharges/discounts for the $\mathbf{B}_t$ bundles by $p_{t\mathbf{J}_t\cup\mathbf{B}_t} = ((p_{tj})_{j\in\mathbf{J}_t}, (p_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{B}_t})$, the price surcharges/discounts for all bundles by $p_{t\mathbf{C}_{t2}} = ((p_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{B}_t}, (p_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_{t2}/\mathbf{B}_t})$, the prices of single products and the price surcharges/discounts for all bundles by $p_{t\mathbf{C}_{t1}} = ((p_{tj})_{j\in\mathbf{J}_t}, (p_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_{t2}})$. Let the market share function in t evaluated at prices $p'_{t\mathbf{C}_{t1}} = ((p'_{tj})_{j\in\mathbf{J}_t}, (p'_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_{t2}})$ and structural parameters $(\delta'_{t\mathbf{J}_t}, \Gamma', \Sigma')$ be $s_t(\delta'_t(\Gamma' + g_t(p'_{t\mathbf{C}_{t2}}; \Sigma'_g)); p'_{t\mathbf{C}_{t1}}, \Sigma'_F)$, the domain of the cost shifters $c_{t\mathbf{J}_t\cup\mathbf{B}_t} = ((c_{tj})_{j\in\mathbf{J}_t}, (c_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{B}_t})$ be $\overline{D}_{tc}$, and remember that $\delta'_{tj} = \Delta'_{tj}(x_{tj}, p_{tj}) + \xi'_{tj}$ for each $j \in \mathbf{J}_t$. Denote the set of equilibrium prices given $\xi_{t\mathbf{J}_t}$ and $c_{t\mathbf{J}_t\cup\mathbf{B}_t}$ by $p_{t\mathbf{J}_t\cup\mathbf{B}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t\cup\mathbf{B}_t}) \subseteq \mathbb{R}^{J_t}_+ \times \mathbb{R}^{B_t}$, given $\xi_{t\mathbf{J}_t}$ by $\overline{\mathbf{P}}_t(\xi_{t\mathbf{J}_t}) = \cup_{c_{t\mathbf{J}_t\cup\mathbf{B}_t}\in\overline{D}_{tc}} p_{t\mathbf{J}_t\cup\mathbf{B}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t\cup\mathbf{B}_t})$, and the grand collection of all possible equilibrium prices by $\overline{\mathbf{P}}_t = \cup_{\xi_{t\mathbf{J}_t}\in D_{t\xi}} \overline{\mathbf{P}}_t(\xi_{t\mathbf{J}_t})$. The vector of observed prices is an equilibrium of the price-setting model, so that $p_{t\mathbf{J}_t\cup\mathbf{B}_t} \in p_{t\mathbf{J}_t\cup\mathbf{B}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t\cup\mathbf{B}_t})$.

With respect to the case of pure components pricing, our identification argument addresses the larger number of endogenous prices by requiring the existence of a proportionally larger number of identifiable cost shifters. In particular, we require the existence of identifiable cost shifters not only for the single products, but also for the bundles in $\mathbf{B}_t$.

Assumption 7.

- *(Cost Shifters at the Bundle-Level)* $\overline{D}_{tc}$ *is open in* $\mathbb{R}^{J_t+B_t}$ *for* $t \in \mathbf{T}$.
- *(Identifiability of Cost Shifters)* $c_{t\mathbf{J}_t\cup\mathbf{B}_t}$ *is a* C^1 *function of* $(\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t\cup\mathbf{B}_t}) \in \{(\xi'_{t\mathbf{J}_t}, p'_{t\mathbf{J}_t\cup\mathbf{B}_t}) : \xi'_{t\mathbf{J}_t} \in D_{t\xi}, p'_{t\mathbf{J}_t\cup\mathbf{B}_t} \in \overline{\mathbf{P}}_t(\xi_{t\mathbf{J}_t})\}$*:* $c_{t\mathbf{J}_t\cup\mathbf{B}_t} = \overline{\phi}_t(\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t\cup\mathbf{B}_t})$.

While Assumption 7 is more demanding than Assumption 4, we believe it to be realistic in many situations. As discussed by Chu et al. (2011), mixed bundling pricing is logistically impractical

[41] Importantly, remember that the demand synergies $\Gamma_{it(1,2)} = \Gamma_t + \zeta_{it(1,2)}$ are heterogeneous across individuals and that only their averages Γ_t are constrained to be common across markets, so that $\Gamma_t = \Gamma$, $t = 1, ..., T$.

for firms with more than a few products, because the number of separate bundles and prices to be monitored and optimized increases exponentially with the number of products. Assumption 7 requires that the act of pricing any bundle of products differently from the sum of the prices of its components entails an additional cost on the side of the firm: in terms of packaging, shelf and storage space, data collection, computational power, surveillance at the cashiers, etc. Similar to Assumption 4 for the case of pure components pricing, it can be shown that Assumption 7 is consistent with commonly employed mixed bundling pricing models: any number of firms (monopoly, duopoly, or oligopoly) playing a complete information simultaneous Bertrand-Nash game with any profile of demand synergies (substitutability and/or complementarity), and a $p_{t\mathbf{J}_t\cup\mathbf{B}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t\cup\mathbf{B}_t})$ of unrestricted cardinality (the price-setting model is allowed to have a unique, several, or infinitely many equilibria).

Define for each $r = 1, ..., R$:

$$\begin{aligned}\overline{\Xi}_r^{\mathrm{ID}} = \{&(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t\cup\mathbf{B}_t})_{t\in\mathbf{T}\setminus\mathbf{T}_0} \in \times_{t\in\mathbf{T}\setminus\mathbf{T}_0}[D_{t\xi}\times\overline{D}_{tc}] : \exists\ t\in\mathbf{T}\setminus\mathbf{T}_0 \text{ such that}\\ &M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(p_{t\mathbf{C}_{t2}};\Sigma_g^0)); p_{t\mathbf{C}_{t1}}, \Sigma_F^0); p_{t\mathbf{C}_{t1}}, \Sigma_F^r) \neq \Gamma^r + g_t(p_{t\mathbf{C}_{t2}};\Sigma_g^r)\\ &\text{for any } p_{t\mathbf{J}_t\cup\mathbf{B}_t} \in p_{t\mathbf{J}_t\cup\mathbf{B}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t\cup\mathbf{B}_t})\}\end{aligned}$$

and $\overline{\Xi}^{\mathrm{ID}} = \cap_{r=1}^R \overline{\Xi}_r^{\mathrm{ID}}$. Similar to Assumption 5 for pure components pricing, we propose the following technical assumption for the case of mixed bundling:

Assumption 8. *For every $r = 1, ..., R$, there exists $t \in \mathbf{T}\setminus\mathbf{T}_0$, so that for almost every $p_{t\mathbf{J}_t\cup\mathbf{B}_t} \in \overline{\mathbf{P}}_t = \cup_{\xi_{t\mathbf{J}_t}\in D_{t\xi}}\overline{\mathbf{P}}_t(\xi_{t\mathbf{J}_t})$, there exists $\xi'_{t\mathbf{J}_t}$, such that $\Gamma^r + g_t(\Sigma_g^r) \neq M_t s_t^{-1}(s_t(\delta'_t(\Gamma^0 + g_t(p_{t\mathbf{C}_{t2}};\Sigma_g^0)); p_{t\mathbf{C}_{t1}}, \Sigma_F^0); p_{t\mathbf{C}_{t1}}, \Sigma_F^r)$.*

As Assumption 5, Assumption 8 also holds given a strengthening of the real analytic property of mixed logit models:

Corollary 3. *Suppose that the following conditions hold:*

- *For $r = 1, ..., R$, $\overline{\Xi}_r^{ID} \neq \emptyset$.*
- *For any $t \in \mathbf{T}\setminus\mathbf{T}_0$ and (Γ, Σ), $s_t(\delta_t(\Gamma + g_t(p_{t\mathbf{C}_{t2}};\Sigma_g)); p_{t\mathbf{C}_{t1}}, \Sigma_F)$ is real analytic with respect to $(\delta_{t\mathbf{J}_t}, p_{t\mathbf{C}_{t1}})$ and $g_t(p_{t\mathbf{C}_{t2}};\Sigma_g)$ is real analytic with respect to $p_{t\mathbf{C}_{t2}}$.*

Then, Assumption 8 holds.

Proof. The proof is similar to that of Corollary 2. □

Denote by $(\xi^0_{t\mathbf{J}_t}, c^0_{t\mathbf{J}_t\cup\mathbf{B}_t}, p^0_{t\mathbf{C}_{t1}})$ the true value of $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t\cup\mathbf{B}_t}, p_{t\mathbf{C}_{t1}})$ for which $s_t(\delta_t(\Gamma^0 + g_t(p^0_{t\mathbf{C}_{t2}};\Sigma_g^0)); p^0_{t\mathbf{C}_{t1}}, \Sigma_F^0) = \mathcal{s}_t$.

Theorem 9. *Suppose Assumptions 1–3 and 7 hold, and Θ_Σ is compact. Then, it follows that:*

- *If* $(\xi^0_{t\mathbf{J}_t}, c^0_{t\mathbf{J}_t\cup\mathbf{B}_t})_{t\in\mathbf{T}\setminus\mathbf{T}_0} \in \overline{\Xi}^{ID}$, *then system (9) has a unique solution in* Θ_Σ *and model (3) is globally identified.*

- *If Assumption 8 holds, the Lebesgue measure of* $\times_{t\in\mathbf{T}\setminus\mathbf{T}_0}[D_{t\xi} \times \overline{D}_{tc}] \setminus \overline{\Xi}^{ID}$ *is zero.*

Proof. Note that given Assumption 7, for any $\mathbf{b} \in \mathbf{C}_{t2}/\mathbf{B}_t$, $p_{t\mathbf{b}} = \sum_{j\in\mathbf{b}} p_{tj}$. Denote this dependence as $p_{t\mathbf{C}_{t2}/\mathbf{B}_t} = p_{t\mathbf{C}_{t2}/\mathbf{B}_t}(p_{t\mathbf{J}_t})$ and then the market share function can be written as:

$$s_t(\delta'_t(\Gamma' + g_t(p_{t\mathbf{C}_{t2}/\mathbf{B}_t}(p_{t\mathbf{J}_t}), p_{t\mathbf{B}_t}; \Sigma'_g)); p_{t\mathbf{J}_t\cup\mathbf{B}_t}, p_{t\mathbf{C}_{t2}/\mathbf{B}_t}(p_{t\mathbf{J}_t}), \Sigma'_F).$$

Then, given (Γ', Σ'), the dependence of s_t on $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t\cup\mathbf{B}_t})$ is channeled through that of $p_{t\mathbf{J}_t\cup\mathbf{B}_t}$ on $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t\cup\mathbf{B}_t})$, as in Theorem 4. To prove Theorem 9, we can then apply the same arguments as in the proof of Theorem 4 on $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t\cup\mathbf{B}_t})$. □

Supplement Tables

Table 8: Average Estimated Own- and Cross-Price Elasticities (Restricted Model 1, $\Gamma_{i\mathbf{b}} = 0$ for each i and $\mathbf{b}$)

		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	−1.247 (0.068)	0.005 (0.000)	0.007 (0.000)	0.003 (0.000)	0.001 (0.000)	0.003 (0.000)	0.001 (0.000)	0.000 (0.000)	0.005 (0.000)	0.001 (0.000)	0.002 (0.000)	0.001 (0.000)	0.002 (0.000)	0.001 (0.000)	0.006 (0.000)	0.001 (0.000)
Kellogg's	R	0.014 (0.001)	−1.435 (0.081)	0.008 (0.001)	0.004 (0.000)	0.001 (0.000)	0.003 (0.000)	0.001 (0.000)	0.000 (0.000)	0.006 (0.000)	0.001 (0.000)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)	0.007 (0.001)	0.001 (0.000)
	F/W	0.016 (0.001)	0.007 (0.000)	−1.197 (0.065)	0.004 (0.000)	0.001 (0.000)	0.004 (0.000)	0.002 (0.000)	0.000 (0.000)	0.008 (0.001)	0.001 (0.000)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)	0.009 (0.001)	0.001 (0.000)
	S	0.020 (0.001)	0.008 (0.001)	0.012 (0.001)	−1.208 (0.065)	0.001 (0.000)	0.004 (0.000)	0.003 (0.000)	0.001 (0.000)	0.010 (0.065)	0.001 (0.000)	0.002 (0.000)	0.003 (0.000)	0.004 (0.000)	0.002 (0.001)	0.011 (0.001)	0.002 (0.000)
Quaker	R	0.013 (0.001)	0.006 (0.000)	0.007 (0.001)	0.003 (0.000)	−1.597 (0.092)	0.003 (0.000)	0.001 (0.000)	.	0.006 (0.000)	0.002 (0.000)	0.003 (0.000)	0.007 (0.000)	0.002 (0.000)	0.002 (0.000)	0.036 (0.002)	0.003 (0.000)
	F/W	0.021 (0.001)	0.008 (0.001)	0.013 (0.001)	0.005 (0.000)	0.001 (0.000)	−1.173 (0.064)	0.002 (0.000)	0.000 (0.000)	0.009 (0.001)	0.001 (0.000)	0.002 (0.000)	0.003 (0.000)	0.003 (0.000)	0.002 (0.000)	0.011 (0.001)	0.001 (0.000)
	S	0.023 (0.002)	0.008 (0.001)	0.014 (0.001)	0.009 (0.001)	0.001 (0.000)	0.005 (0.000)	−1.259 (0.069)	0.001 (0.000)	0.011 (0.001)	0.001 (0.000)	0.003 (0.000)	0.002 (0.000)	0.005 (0.000)	0.003 (0.000)	0.013 (0.001)	0.002 (0.000)
Post	R	0.006 (0.000)	0.002 (0.000)	0.004 (0.000)	0.011 (0.001)	.	0.003 (0.000)	0.002 (0.000)	−0.963 (0.051)	0.002 (0.000)	.	0.004 (0.000)	0.000 (0.000)	.	0.004 (0.000)	0.022 (0.001)	0.007 (0.000)
	F/W	0.017 (0.001)	0.007 (0.000)	0.011 (0.001)	0.005 (0.000)	0.001 (0.000)	0.003 (0.000)	0.002 (0.000)	0.000 (0.000)	−1.201 (0.065)	0.001 (0.000)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)	0.009 (0.001)	0.001 (0.000)
	S	0.031 (0.002)	0.010 (0.001)	0.015 (0.001)	0.013 (0.001)	0.003 (0.000)	0.006 (0.000)	0.003 (0.000)	.	0.017 (0.001)	−1.158 (0.063)	0.004 (0.000)	0.002 (0.000)	0.005 (0.000)	0.010 (0.001)	0.017 (0.001)	0.003 (0.000)
Private	R	0.014 (0.001)	0.006 (0.000)	0.008 (0.001)	0.004 (0.000)	0.001 (0.000)	0.003 (0.000)	0.001 (0.000)	0.001 (0.000)	0.007 (0.000)	0.001 (0.000)	−1.015 (0.054)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)	0.007 (0.000)	0.001 (0.000)
	F/W	0.018 (0.001)	0.008 (0.001)	0.012 (0.001)	0.005 (0.000)	0.005 (0.000)	0.004 (0.000)	0.002 (0.000)	0.000 (0.000)	0.008 (0.001)	0.001 (0.000)	0.002 (0.000)	−0.944 (0.049)	0.002 (0.000)	0.002 (0.000)	0.009 (0.001)	0.001 (0.000)
	S	0.020 (0.001)	0.009 (0.001)	0.013 (0.001)	0.008 (0.001)	0.001 (0.000)	0.004 (0.000)	0.003 (0.000)	.	0.009 (0.001)	0.001 (0.000)	0.003 (0.000)	0.002 (0.000)	−0.945 (0.049)	0.002 (0.000)	0.011 (0.001)	0.002 (0.000)
SmallP.	R	0.016 (0.001)	0.007 (0.000)	0.009 (0.001)	0.004 (0.000)	0.001 (0.000)	0.003 (0.000)	0.001 (0.000)	0.001 (0.000)	0.007 (0.000)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)	−1.636 (0.091)	0.008 (0.001)	0.001 (0.000)
	F/W	0.017 (0.001)	0.007 (0.001)	0.011 (0.001)	0.005 (0.000)	0.005 (0.000)	0.004 (0.000)	0.002 (0.000)	0.001 (0.000)	0.008 (0.001)	0.001 (0.000)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)	−1.370 (0.076)	0.001 (0.000)
	S	0.023 (0.002)	0.009 (0.001)	0.014 (0.001)	0.010 (0.001)	0.004 (0.000)	0.005 (0.000)	0.003 (0.000)	0.001 (0.000)	0.011 (0.001)	0.002 (0.000)	0.003 (0.000)	0.001 (0.000)	0.005 (0.000)	0.002 (0.000)	0.013 (0.001)	−0.994 (0.053)

Notes: The Table reports the average estimated own- (diagonal) and cross-price (off-diagonal) elasticities from restricted model 1 (which constrains $\Gamma_{i\mathbf{b}} = 0$ in estimation), where averages are computed across markets. Each entry reports the percent change in the brand-level 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 brand-level 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." The missing values "·" refer to situations where the corresponding column and row brands are never simultaneously sold in the same market. The standard errors of the estimated elasticities are in brackets obtained from a parametric bootstrap as in Nevo (2000, 2001) with 50 draws.

Table 9: Average Estimated Own- and Cross-Price Elasticities (Restricted Model 2, $\Gamma_{i\mathbf{b}} = -\infty$ for each i and $\mathbf{b}$)

		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	−4.027 (0.699)	0.094 (0.018)	0.135 (0.025)	0.049 (0.009)	0.014 (0.003)	0.039 (0.007)	0.015 (0.003)	0.003 (0.001)	0.095 (0.018)	0.007 (0.001)	0.026 (0.005)	0.023 (0.004)	0.022 (0.004)	0.025 (0.005)	0.109 (0.020)	0.010 (0.002)
Kellogg's	R	0.276 (0.052)	−4.716 (0.838)	0.117 (0.022)	0.042 (0.008)	0.014 (0.003)	0.030 (0.006)	0.014 (0.003)	0.003 (0.001)	0.085 (0.016)	0.006 (0.001)	0.027 (0.005)	0.020 (0.004)	0.019 (0.004)	0.026 (0.005)	0.097 (0.018)	0.008 (0.002)
	F/W	0.323 (0.060)	0.093 (0.017)	−3.869 (0.670)	0.054 (0.010)	0.014 (0.003)	0.040 (0.007)	0.017 (0.003)	0.004 (0.001)	0.109 (0.020)	0.007 (0.001)	0.026 (0.005)	0.025 (0.005)	0.023 (0.004)	0.024 (0.004)	0.115 (0.021)	0.009 (0.002)
	S	0.329 (0.060)	0.098 (0.018)	0.154 (0.028)	−3.868 (0.667)	0.010 (0.002)	0.038 (0.007)	0.027 (0.005)	0.004 (0.001)	0.120 (0.022)	0.011 (0.002)	0.026 (0.005)	0.025 (0.004)	0.040 (0.007)	0.023 (0.004)	0.121 (0.023)	0.016 (0.003)
Quaker	R	0.276 (0.051)	0.100 (0.019)	0.102 (0.019)	0.040 (0.007)	−5.253 (0.945)	0.033 (0.006)	0.013 (0.002)	.	0.084 (0.016)	0.005 (0.001)	0.037 (0.007)	0.017 (0.003)	0.020 (0.004)	0.026 (0.005)	0.096 (0.018)	0.009 (0.002)
	F/W	0.349 (0.065)	0.092 (0.017)	0.152 (0.028)	0.050 (0.009)	0.014 (0.003)	−3.781 (0.655)	0.018 (0.003)	0.004 (0.001)	0.101 (0.019)	0.007 (0.001)	0.028 (0.005)	0.029 (0.005)	0.026 (0.005)	0.026 (0.005)	0.122 (0.023)	0.010 (0.002)
	S	0.336 (0.062)	0.093 (0.017)	0.151 (0.028)	0.085 (0.016)	0.016 (0.003)	0.046 (0.008)	−4.053 (0.704)	0.004 (0.001)	0.115 (0.021)	0.010 (0.002)	0.028 (0.005)	0.024 (0.004)	0.041 (0.008)	0.026 (0.005)	0.130 (0.025)	0.016 (0.003)
Post	R	0.212 (0.038)	0.046 (0.009)	0.101 (0.018)	0.030 (0.005)	.	0.052 (0.010)	0.005 (0.001)	−2.961 (0.506)	0.048 (0.009)	.	0.012 (0.002)	0.008 (0.001)	.	0.011 (0.002)	0.063 (0.011)	0.019 (0.003)
	F/W	0.314 (0.058)	0.095 (0.018)	0.151 (0.028)	0.057 (0.010)	0.014 (0.003)	0.037 (0.007)	0.018 (0.003)	0.004 (0.001)	−3.849 (0.665)	0.007 (0.001)	0.029 (0.005)	0.025 (0.005)	0.022 (0.004)	0.023 (0.004)	0.112 (0.021)	0.011 (0.002)
	S	0.364 (0.067)	0.093 (0.017)	0.143 (0.026)	0.105 (0.019)	0.009 (0.002)	0.046 (0.008)	0.024 (0.004)	.	0.164 (0.029)	−3.678 (0.633)	0.032 (0.006)	0.016 (0.003)	0.040 (0.007)	0.024 (0.004)	0.149 (0.027)	0.021 (0.004)
Private	R	0.264 (0.048)	0.092 (0.017)	0.110 (0.020)	0.039 (0.007)	0.016 (0.003)	0.032 (0.006)	0.013 (0.002)	0.004 (0.001)	0.088 (0.016)	0.005 (0.001)	−3.179 (0.544)	0.021 (0.004)	0.019 (0.003)	0.025 (0.005)	0.090 (0.017)	0.007 (0.001)
	F/W	0.306 (0.056)	0.089 (0.016)	0.142 (0.026)	0.050 (0.009)	0.013 (0.002)	0.043 (0.008)	0.016 (0.003)	0.004 (0.001)	0.103 (0.018)	0.006 (0.001)	0.028 (0.005)	−2.928 (0.494)	0.023 (0.004)	0.023 (0.004)	0.102 (0.019)	0.010 (0.002)
	S	0.325 (0.059)	0.095 (0.017)	0.144 (0.026)	0.088 (0.016)	0.013 (0.002)	0.043 (0.008)	0.027 (0.005)	.	0.099 (0.018)	0.010 (0.002)	0.028 (0.005)	0.024 (0.004)	−2.925 (0.491)	0.022 (0.004)	0.113 (0.021)	0.015 (0.003)
Small P.	R	0.291 (0.055)	0.101 (0.019)	0.115 (0.021)	0.040 (0.007)	0.012 (0.002)	0.034 (0.006)	0.014 (0.003)	0.003 (0.001)	0.083 (0.015)	0.005 (0.001)	0.029 (0.005)	0.020 (0.004)	0.019 (0.003)	−5.300 (0.942)	0.099 (0.019)	0.009 (0.002)
	F/W	0.325 (0.060)	0.097 (0.018)	0.144 (0.027)	0.053 (0.010)	0.014 (0.003)	0.041 (0.008)	0.018 (0.003)	0.003 (0.001)	0.101 (0.019)	0.008 (0.001)	0.027 (0.005)	0.023 (0.004)	0.023 (0.004)	0.025 (0.005)	−4.479 (0.785)	0.011 (0.002)
	S	0.324 (0.059)	0.098 (0.018)	0.147 (0.027)	0.089 (0.016)	0.011 (0.002)	0.041 (0.008)	0.027 (0.005)	0.004 (0.001)	0.109 (0.020)	0.012 (0.002)	0.029 (0.005)	0.023 (0.004)	0.041 (0.008)	0.022 (0.004)	0.126 (0.023)	−3.111 (0.527)

Notes: The Table reports the average estimated own- (diagonal) and cross-price (off-diagonal) elasticities from restricted model 2 (which constrains $\Gamma_{i\mathbf{b}} = -\infty$ in estimation), where averages are computed across markets. Each entry reports the percent change in the market share of the column RTE cereal brand with respect to a 1% increase in the price of the row RTE cereal brand. Cereal type R refers to "Regular," F/W to "Fiber/Whole Grain," and S to "Added Sugar." The missing values "·" refer to situations where the corresponding column and row brands are never simultaneously sold in the same market. The standard errors of the estimated elasticities are in brackets obtained from a parametric bootstrap as in Nevo (2000, 2001) with 50 draws.