Econometric Theory, **0**, 2024, 1–49.
doi:10.1017/S0266466624000148

REAL ANALYTIC DISCRETE CHOICE MODELS OF DEMAND: THEORY AND IMPLICATIONS

ALESSANDRO IARIA
University of Bristol and CEPR

AO WANG
University of Warwick and CAGE Research Centre

We demonstrate that a large class of discrete choice models of demand can be approximated by real analytic demand models. We obtain this result by combining (i) a novel real analytic property of the mixed logit and the mixed probit models with any distribution of random coefficients and (ii) an approximation property of finite mixtures of Gumbel and Gaussian distributions. To illustrate some of the implications of this result, we discuss how real analyticity facilitates nonparametric and semi-nonparametric identification, extrapolation to hypothetical counterfactuals, numerical implementation of demand inverses, and numerical implementation of the maximum likelihood estimator.

1. INTRODUCTION

Real analyticity is an extreme form of smoothness of a function's dependence on its arguments: a real analytic function $f(x)$ is an infinitely differentiable function whose Taylor series at any point x_0 in its domain converges to $f(x)$ for x in a neighborhood of x_0 (Rudin, 1976). Intuitively, a real analytic function can be represented as a power series (i.e., an infinite degree polynomial) and manipulated in the same way as polynomials within an open interval of convergence. In the context of discrete choice models of demand, the restrictions imposed by real analyticity can limit the realism of a demand model (e.g., ruling out kinks and discontinuities) but, when economically affordable, they can also facilitate econometric implementation. While some recent papers have relied on the high-level use of real analyticity for specific identification (Fox et al., 2012; Fox and Gandhi, 2016; Allen and Rehbeck, 2020; Wang, 2023) and estimation (Wang, 2023) results, it is unclear how broad the class of real analytic discrete choice

We are grateful to the Editor (Peter C. B. Phillips), the Co-Editor (Simon Lee), and two anonymous referees for comments and suggestions which greatly improved the article. An early version of this paper circulated under the title "*The Mixed Logit and Mixed Probit are Real Analytic*." Address correspondence to Ao Wang, Department of Economics, University of Warwick and CAGE Research Centre, Coventry, United Kingdom, e-mail: ao.wang@warwick.ac.uk

models of demand is and in which sense the restrictions imposed by real analyticity can be econometrically and/or numerically advantageous.

We consider a class of discrete choice models of demand with an index structure in the indirect utilities (possibly nonlinear), with any distribution of random coefficients, and—in line with Berry, Levinsohn, and Pakes (1995)—which can include endogenous regressors. In the first part of the paper, we demonstrate that this class of models has a dense real analytic subset: any demand model in this class (which needs *not* be real analytic) can be approximated uniformly and arbitrarily well by a real analytic demand model from the same class. We obtain this result by showing that the mixed logit (McFadden and Train, 2000) and the mixed probit (Hausman and Wise, 1978) are real analytic for any distribution of random coefficients, and then by relying on an approximation property of finite mixtures of Gumbel and Gaussian distributions (Nguyen et al., 2020).

In prior work, Fox et al. (2012) showed that the mixed logit model with random coefficients defined over a compact support is real analytic. This compact support condition however rules out distributions of random coefficients such as the normal and the log-normal, which are commonly used by applied researchers. We generalize this result and demonstrate that the mixed logit is real analytic for any distribution of random coefficients. Differently, we are not aware of prior work showing that the mixed probit is real analytic: for example, Stinchcombe and White (1998) state (without proof) that the normal cumulative distribution function is not supposed to be real analytic (first sentence after Thm. 3.10, p. 305). We instead prove that also the mixed probit model is real analytic for any distribution of random coefficients. In addition, we further show that any discrete choice model of demand belonging to our class (not necessarily a mixed logit or a mixed probit) can be approximated uniformly and arbitrarily well by a real analytic finite mixture of mixed logits and/or mixed probits.

The class of discrete choice models of demand we consider is economically "broad" in the sense of McFadden and Train (2000). In their famous "possibility result," McFadden and Train (2000, Thm. 1, p. 451) showed that any random utility model within a large class can be approximated by a flexible mixed logit model belonging to the class we consider. Our result then implies that one can approximate uniformly and arbitrarily well any such mixed logit approximant by a real analytic demand model, therefore uniformly approximating any random utility model in the class considered by McFadden and Train (2000). In other words, when dealing with any random utility model in the class considered by McFadden and Train (2000), one can restrict attention to a subset of real analytic demand models without any loss of precision.

In the second part of the paper, we discuss the econometric advantages of real analytic demand models in terms of nonparametric and semi-nonparametric identification, extrapolation to hypothetical counterfactuals, numerical implementation of demand inverses in the context of aggregate market-level data, and numerical implementation of the maximum likelihood estimator (MLE) in the context of disaggregate individual-level data.

In order to reduce the potential for misspecification, researchers are often interested in the nonparametric identification of discrete choice models of demand (Berry and Haile, 2014, 2018). This allows demand models to be robust with respect to a broad range of consumer preferences and behaviors, including complex substitution patterns, continuous quantity choices, complementarities across products, and consumer inattention, and consequently it allows researchers to perform more realistic counterfactual simulations, such as hypothetical mergers or the introduction of new taxes (Compiani, 2022). Without further restrictions, nonparametrically identified demand models can only be used to predict counterfactual outcomes within the support spanned by the data. However, it is possible for counterfactual outcomes of interest to fall outside the support of the data, for example, equilibrium prices that significantly increase after a hypothetical merger. Identifying such counterfactual outcomes is an extrapolation exercise for which nonparametric identification may not be sufficient without additional restrictions on the demand model. We show that real analyticity implies a form of identification extensibility that overcomes this challenge.

Real analyticity also facilitates the semi-nonparametric identification of discrete choice models of demand. In this approach, the researcher assumes, for example, a mixed logit or a mixed probit model, and then wishes to identify the distribution of random coefficients nonparametrically. Without real analyticity, the semi-nonparametric identification of demand models usually requires covariates with large support (restrictive for example with price variables, known to be positive) and prevents the inclusion of interactions among covariates (Fox et al., 2012; Fox and Gandhi, 2016; Masten, 2018). These standard assumptions, in some cases, limit the economic content of discrete choice models of demand. For example, the approximation result by McFadden and Train (2000) crucially relies on the inclusion of interactions among covariates (see p. 466). The identification results by Fox et al. (2012), however, explicitly rule out interactions among covariates (Assumptions 2 and 3, pp. 207–208). Differently, by generalizing the mixed logit identification result in Wang (2023) to real analytic discrete choice models of demand, we show that the distribution of random coefficients is nonparametrically identified also in the presence of interactions among covariates.

In the context of aggregate market-level data, both the parametric BLP (Berry et al., 1995) and the semi-nonparametric approach (Wang, 2023) require the computation of a demand inverse for each market and at each iteration of the generalized method of moments (GMM) minimization. Inversion can be numerically challenging (Knittel and Metaxoglou, 2014), motivating different numerical approaches such as fixed point algorithms (Berry et al., 1995; Lee and Seo, 2016), the Mathematical Program with Equilibrium Constraints (MPEC) (Dubé, Fox, and Su, 2012a; Su and Judd, 2012), and the use of an approximate inverse (Salanié and Wolak, 2019). The real analyticity of the demand model mitigates this challenge by guaranteeing the desirable numerical performance of Newton–Raphson (NR) algorithms.

As is well known, the NR algorithm can achieve quadratic convergence to the unique solution giving rise to the demand inverse when the starting value of the iterations is "close" to the unique solution and the demand function is twice continuously differentiable (Lee and Seo, 2016). However, the extent of the proximity of the starting value to the unique solution typically depends on knowledge of the demand model that is not available to the researcher before estimation. In addition, numerical convergence is usually determined by the researcher, who must choose some tolerance within which to stop the iterations. Unfortunately, though, there is little theoretical guidance on how to choose such level of tolerance, which is often calibrated using heuristic rules of thumb. In fact, despite achieving numerical convergence on the basis of such rules of thumb, it is still possible for the NR algorithm to be divergent. Unguided choices of starting values and of stopping criteria represent a challenge for the numerical convergence of NR algorithms to the unique solution giving rise to the demand inverse (Dubé et al., 2012a; Lee and Seo, 2016; Conlon and Gortmaker, 2020).

Building on Smale's alpha-theory (Smale, 1986), we show that the real analyticity of the demand model leads to simple and verifiable sufficient conditions that do not require prior knowledge of the parameters one is trying to estimate and that guarantee the supra-exponential convergence of the NR algorithm to the unique solution giving rise to the demand inverse.[1] These verifiable sufficient conditions are simple to compute, do not require the implementation of any NR iteration, and provide theoretical guidance for the selection of robust starting values and stopping criteria that guarantee the quick convergence of the NR algorithm. In Monte Carlo simulations, we investigate the practical performance of these sufficient conditions both when used in stand-alone NR algorithms and when embedded in a hybrid algorithm (HA) that combines FP and NR iterations (Rust, 1987; Iskhakov et al., 2016). We show that the proposed HA is guaranteed to converge from any starting value and that its numerical performance in the context of demand inverses is superior to those of both stand-alone FP and stand-alone NR algorithms.[2]

In the context of demand estimation with disaggregate individual-level data, researchers often rely on (parametric) MLE (Goolsbee and Petrin, 2004; Train and Winston, 2007; Dubois, Griffith, and O'Connell, 2020). For reasons similar to those discussed above, the real analyticity of the demand model (and so of the log-likelihood function) also guarantees the desirable numerical performance of NR algorithms for the implementation of the MLE. The practical performance of NR algorithms to implement the MLE is subject to the same challenges mentioned above: the appropriate choices of starting values and of stopping criteria often depend on knowledge of the demand model one is trying to estimate. As for the computation of demand inverses, we show that the real analyticity of the log-likelihood function gives rise to simple and verifiable sufficient conditions for the supra-exponential convergence of NR algorithms to a local maximum of the

[1] As shown in Rheinboldt (1988), the supra-exponential rate is of the same order as the rate of quadratic convergence.

[2] Stand-alone FP algorithms based on contraction mappings are guaranteed to converge to the unique solution from any starting value, but at a slower rate than the proposed HA. Stand-alone NR algorithms may instead fail to converge from starting values that are not in the basin of attraction of the unique solution.

log-likelihood function that do not require any prior knowledge of the parameters one is trying to estimate. By speeding up the computation of all local maxima of the log-likelihood function, these sufficient conditions also prove useful in mitigating the consequences of multiplicity of local maxima in the numerical implementation of the MLE.

Some caution is needed in interpreting the econometric implications of real analyticity discussed in the second part of the paper, as they apply directly only to the real analytic demand models and to the real analytic approximants described in the first part of the paper, but not necessarily to more general demand models that are *not* real analytic, for example, demand models that exhibit discontinuities and kinks (Smith, 1935). While in this paper, we show that a large class of these more general demand models can be approximated uniformly and arbitrarily well by real analytic demand models, it is not known whether some form of "transitivity" holds between the real analytic approximants and the targets of approximation, and so whether any of the econometric implications of real analyticity also hold for the approximated models.

The rest of the paper is organized as follows: Section 2 introduces the notation and the class of discrete choice models of demand we consider. Section 3 defines the concept of real analyticity and presents the main results of the paper. Section 4 discusses the implications of real analyticity for nonparametric and semi-nonparametric identification. Section 5 discusses the implications of real analyticity for the numerical implementation of demand inverses and of the MLE. Section 6 reports some Monte Carlo simulations related to the results from Section 5. Section 7 concludes the paper. All the proofs, intermediate results, and details about the Monte Carlo simulations are reported in the Appendix. The MATLAB code used to perform the Monte Carlo simulations is available on the authors' webpages.

2. A CLASS OF DEMAND MODELS

In this section, we describe the class of discrete choice models of demand studied in the paper.

Each individual i in cell t (e.g., a time period, a market, or a combination of both) is observed to choose an alternative from choice set $\{0,1,\dots,J\}$, where 0 denotes the outside option. Individual i's indirect utility from choosing alternative $j=1,\dots,J$ in t is

$$U_{ijt} = g_j(x_{jt}, d_i; \beta_i) + \xi_{jt} + \varepsilon_{ijt}, \tag{1}$$

where g_j is a known alternative-specific function of j's observed K-dimensional characteristics x_{jt} (e.g., price), individual i's observed demographics d_i (e.g., gender, income, and education), and a finite-dimensional vector of individual-specific random coefficients β_i (e.g., price coefficient); ξ_{jt} is an unobserved (j,t)-specific intercept common to all individuals in cell t (e.g., demand shock at the product-market level), and ε_{ijt} is an idiosyncratic error term. Individual i's indirect utility from choosing the outside option 0 in t is normalized to ε_{i0t}:

$$U_{i0t} = \varepsilon_{i0t}.$$

Remark 1. Our results apply to both cross-sectional and panel data. Each individual can either be observed only once making a choice from a specific cell t (cross-sectional data) or repeatedly making choices over several t's (panel data).

Remark 2. We place no restriction on (ξ_{jt}, x_{jt}, d_i), which can be of any type (e.g., discrete or continuous, with unbounded or bounded support) and, in line with Berry et al. (1995), allow for the possibility of correlation between ξ_{jt} and x_{jt} (e.g., price endogeneity).

Remark 3. The function $g_j(x_{jt}, d_i; \beta_i)$ does not impose any restriction on the interactions among its arguments and only needs to be known up to the finite-dimensional vector of random coefficients β_i. For example, as required by the approximation result in McFadden and Train (2000), one could use sieves or polynomials to specify a flexible linear index $g_j(x_{jt}, d_i; \beta_i) = \mathcal{X}_j(x_{jt}, d_i)\beta_i$ with $\mathcal{X}_j(x_{jt}, d_i)$ being a known vector of interactions among the elements of (x_{jt}, d_i).

We assume that in each cell t, the idiosyncratic errors $(\varepsilon_{ijt})_{j=0}^J$ are distributed according to G and independently of all other components of $(U_{ijt})_{j=0}^J$. Moreover, conditional on the observed demographics d_i, β_i is distributed according to $F(\cdot; d_i)$ and independently of $(\xi_{jt}, x_{jt})_{j=1}^J$ and $(\varepsilon_{ijt})_{j=0}^J$. Then, individual i's choice probability of j in t is

$$\begin{aligned} p_{ijt} &= \sigma_j(\xi_t; X_t, F, d_i, G) \\ &= \int \mathbf{1}\{U_{ijt} > U_{ijr}, \forall r \neq j\} dF(\beta_i; d_i) dG(\varepsilon_{i0t}, \ldots, \varepsilon_{iJt}), \end{aligned} \quad \textbf{(2)}$$

where $\xi_t = (\xi_{jt})_{j=1}^J$, $X_t = (x_{jt})_{j=1}^J$, and $\sigma_j(\cdot)$ is the choice probability function of j.

Discrete choice model (2) can be used in most settings of interest for applied researchers, including disaggregate individual-level data (Goolsbee and Petrin, 2004; Train and Winston, 2007; Dubois et al., 2020), aggregate market-level data (Berry et al., 1995; Petrin, 2002; Wang, 2023), and combinations of both (Berry, Levinsohn, and Pakes, 2004). With aggregate market-level data, the researcher only observes market shares $(p_{jt})_{j=1}^J$, rather than individual choices, and the distribution of demographics within each market t, $\Pi_t(d_i)$. In this case, $\sigma_j(\xi_t; X_t, F, d_i, G)$ is typically further integrated over d_i, so that $p_{jt} = \int \sigma_j(\xi_t; X_t, F, d_i, G) d\Pi_t(d_i)$.

3. REAL ANALYTICITY

In this section, we first define the concept of real analyticity. We then show that when the distribution G of the $J+1$ idiosyncratic errors $(\varepsilon_{ijt})_{j=0}^J$ is either i.i.d. Gumbel or nondegenerate multivariate Gaussian, the resulting mixed logit or mixed probit model in (2) is real analytic. Finally, we state our main real analytic

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

approximation result: *any* demand model (2) (i.e., for any distributions F and G) can be approximated arbitrarily well by a real analytic finite mixture of mixed logit and/or mixed probit models as in (2).

We use $\|\cdot\|$ to refer to the Euclidean norm and $\|\cdot\|_{L_p}$ to the L_p norm of a function for $p > 0$. Moreover, we use $\mathcal{F}$ to denote the set of all possible distributions F.

Definition: Real Analyticity.

We define $\sigma_j(\xi_t; X_t, F, d_i, G), j = 1, \ldots, J$, to be real analytic with respect to ξ_t at ξ_{t0} (and with respect to X_t at X_{t0}) if:

1. $\sigma_j(\xi_t; X_t, F, d_i, G)$ is infinitely differentiable with respect to ξ_t at ξ_{t0} (and with respect to X_t at X_{t0}).
2. There exists an open neighborhood of ξ_{t0} (and X_{t0}) such that the Taylor series of $\sigma_j(\xi_t; X_t, F, d_i, G)$ at ξ_{t0} (and X_{t0}) converges to $\sigma_j(\xi_t; X_t, F, d_i, G)$ for any ξ_t (and X_t) in this neighborhood.[3]

When a function is real analytic with respect to one of its arguments at each point in the corresponding domain, we say that the function is real analytic with respect to this argument in the domain. The first requirement holds whenever the distribution G is sufficiently smooth. Violations happen, for example, when the distribution function G exhibits jumps, so that the resulting demand function is discontinuous, or again when G has kinks, so that the demand function is non-differentiable (even though potentially continuous).[4] Despite these important exceptions, the first requirement is trivially satisfied by both the mixed logit and the mixed probit models. In contrast, even for these two relatively simple models, it is not trivial to verify whether the second requirement holds, mainly because of the unconstrained distribution of random coefficients F we consider.

To understand the nature of this challenge, suppose G is i.i.d. Gumbel, so that model (2) is a mixed logit. Denote by Supp(F) the support of F. The simplest possible case occurs when $\text{Supp}(F) = \{\beta\}$, that is, $\Pr(\beta_i = \beta) = 1$. Then, (2) further simplifies to a logit model:

$$p_{ijt} = \sigma_j(\xi_t; X_t, F, d_i, G) = \frac{\exp\{\xi_{jt} + g_j(x_{jt}, d_i; \beta)\}}{1 + \sum_{j'=1}^{J} \exp\{\xi_{j't} + g_r(x_{j't}, d_i; \beta)\}}. \quad (3)$$

Because $\exp\{\xi_{j't} + g_r(x_{j't}, d_i; \beta)\}$ is real analytic with respect to $\xi_{j't} \in \mathbb{R}$ for $j' = 1, \ldots, J$, then the real analyticity of $\sigma_j(\xi_t; X_t, F, d_i, G)$ holds trivially. Similarly, when F has finite support, say $\text{Supp}(F) = \{\beta^1, \ldots, \beta^M\}$, (2) is a mixture of M logit models each with $\beta_i = \beta^m$, $m = 1, \ldots, M$. Because each logit model is

[3] Following Rudin's (1976) Theorem 8.4 (p. 176), this definition of real analyticity is equivalent to Rudin's (1976) definition based on power series (p. 172).

[4] Smith (1935) provides an example of a discontinuous demand function, where the discontinuities are generated by the presence of marginal buyers.

real analytic and has positive radius of convergence at any $\xi_t \in \mathbb{R}^J$, denoted by $r_m(\xi_t; G) > 0$, then the finite mixture also has positive radius of convergence at ξ_t, $r(\xi_t; F, G) := \inf_{m=1,\dots,M} r_m(\xi_t; G) > 0$. For $\|\xi_t' - \xi_t\| < r(\xi_t; G)$, the Taylor series of $\sigma_j(\xi_t'; X_t, F, d_i, G)$ at ξ_t converges to $\sigma_j(\xi_t'; X_t, F, d_i, G)$ and the second requirement is thus satisfied.

When $\text{Supp}(F)$ is bounded, that is, β_i is bounded, intuitively, $r_m(\xi_t; F, G)$ is a continuous function of m in the closure of $\text{Supp}(F)$. Because $r_m(\xi_t; F, G)$ is always positive, then its minimum in the closure of $\text{Supp}(F)$ is guaranteed to be positive and therefore $r(\xi_t; F, G) = \inf_{m \in \text{Supp}(F)} r_m(\xi_t; G) > 0$. As a result, the mixed logit with bounded random coefficients is also real analytic.

However, when $\text{Supp}(F)$ is unbounded, for example, F is Gaussian with $\text{Supp}(F) = \mathbb{R}^J$ or log-normal with $\text{Supp}(F) = \mathbb{R}^J_+$, the argument above does not apply. Even though each logit model in the mixture has positive radius of convergence $r_m(\xi_t; F) > 0$, it is not straightforward to see that their infimum in $\text{Supp}(F)$, $r(\xi_t; F, G)$, is still positive. The following proposition deals with this challenge by showing that $r(\xi_t; F, G)$ can be uniformly bounded away from zero by a constant that depends neither on $\xi_t \in \mathbb{R}^J$ nor on the support of F, so that the radius of convergence of both the mixed logit and the mixed probit with any distribution F is always positive at any ξ_t.

THEOREM 1. *For $j = 1, \dots, J$, $\sigma_j(\xi_t; X_t, F, d_i; G)$ is real analytic with respect to $\xi_t \in \mathbb{R}^J$ and $\inf_{\xi_t \in \mathbb{R}^J, F \in \mathcal{F}} r(\xi_t; F, G) > 0$ when:*

(a) *(mixed logit) G is i.i.d. Gumbel, or*
(b) *(mixed probit) G is multivariate Gaussian.*

In particular, $\inf_{\xi_t \in \mathbb{R}^J, F \in \mathcal{F}} r(\xi_t; F, G) = +\infty$ when G is multivariate Gaussian.

Theorem 1(a) highlights that the radius of convergence of mixed logit (2) for given F at ξ_t depends neither on ξ_t nor on F. Theorem 1(b) shows that the real analyticity of mixed probit (2) is even stronger, with the radius of convergence being infinity. This stronger result is a consequence of tighter bounds on the higher-order derivatives of the mixed probit function with respect to ξ_t, which alleviate the requirement on the size of the radius of convergence.[5]

3.1. Real Analytic Approximations

In this section, we build on Theorem 1 to construct a set of real analytic demand models (2) that combine mixed logit and mixed probit models and that can be used to approximate arbitrarily well *any* demand model (2), that is, not necessarily a real analytic demand model.

[5] We thank an anonymous referee for pointing out the importance of these bounds and their potential usefulness in strengthening several results in this paper.

We start by defining a set of demand models (2) to which we can extend the real analyticity from Theorem 1 and that will serve as approximants. First, we define a set of density functions generated by finite mixtures of density h: $\mathbb{R}^{J+1} \to \mathbb{R}^+$:

$$\mathcal{M}^h = \left\{ \sum_{i=1}^m c_i \frac{1}{\varsigma_i^{J+1}} h\left(\frac{(\varepsilon_j)_{j=0}^J - \mu_i}{\varsigma_i} \right),\ \sum_{i=1}^m c_i = 1,\ c_i > 0,\ \mu_i \in \mathbb{R}^{J+1},\ \varsigma_i > 0,\ 1 \le i \le m; m \in \mathbb{N} \right\}. \tag{4}$$

When $h = \phi$ is a Gaussian density, $\mathcal{M}^\phi$ defines the family of Gaussian mixtures generated by ϕ. When $h = \psi$ is the i.i.d. Gumbel density, $\mathcal{M}^\psi$ defines the family of i.i.d. Gumbel mixtures generated by ψ. We also define a more general family of mixture distributions which includes both $\mathcal{M}^\phi$ and $\mathcal{M}^\psi$:

$$\mathcal{M}^\phi + \mathcal{M}^\psi = \{ rg_1 + (1-r)g_2,\ r \in [0,1],\ g_1 \in \mathcal{M}^\phi,\ g_2 \in \mathcal{M}^\psi \}. \tag{5}$$

To simplify exposition, we use the notation $G \in \mathcal{M}^h$ or $F \in \mathcal{M}^h$ to refer to the fact that the density function corresponding to G or to F belongs to $\mathcal{M}^h$.

Second, we define a family of distributions $\mathcal{F}^e$ whose absolute moments can increase at most at an exponential rate:

$$\mathcal{F}^e = \{ F : \exists A > 1 \text{ such that any } \alpha = (\alpha_1, \ldots, \alpha_K)\text{th absolute moment of } F,\ m_\alpha^F \le A^{\sum_{k=1}^K \alpha_k} \}, \tag{6}$$

where $m_\alpha^F := \int \prod_{k=1}^K |\beta_k|^{\alpha_k} dF(\beta_1, \ldots, \beta_K)$. Leading examples in $\mathcal{F}^e$ are distributions with bounded support, that is, there exists some $A > 1$ such that $|\beta_{ik}| \le A$ almost surely for $k = 1, \ldots, K$. Similarly, we define a family of distributions whose absolute moments can increase at a rate that is the product of an exponential and the squared root of a factorial rate:

$$\mathcal{F}^{e+} = \Bigg\{ F : \exists A > 1 \text{ such that any } \alpha = (\alpha_1, \ldots, \alpha_K)\text{th absolute moment of } F,$$
$$m_\alpha^F \le A^{\sum_{k=1}^K \alpha_k} \sqrt{\prod_{k=1}^K \alpha_k!} \Bigg\}. \tag{7}$$

It is clear that $\mathcal{F}^e \subset \mathcal{F}^{e+}$. Because the moments of the Gaussian distribution increase at a double factorial rate, the multivariate Gaussian distribution with $\beta_{ik} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ belongs to $\mathcal{F}^{e+}$ (but not to $\mathcal{F}^e$).[6] Finally, we define the set of density

[6]To see this, denote by Φ the distribution function of the multivariate Gaussian with $\beta_{ik} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. Note that the $\alpha = (\alpha_1, \ldots, \alpha_K)$th absolute moment of Φ is $m_\alpha^\Phi = \prod_{k=1}^K (\alpha_k!!) \left(\mathbf{1}\{\alpha_k \text{ is odd}\} \sqrt{2/\pi} + \mathbf{1}\{\alpha_k \text{ is even}\} \right)$, where the double factorial $\alpha_k!! = (\alpha_k)(\alpha_k - 2)(\alpha_k - 4) \ldots 2$ if α_k is even or $\alpha_k!! = (\alpha_k)(\alpha_k - 2)(\alpha_k - 4) \ldots 1$ if α_k is odd. For any even α_k, $\alpha_k! = \alpha_k!! \times (\alpha_k - 1)!! = (\alpha_k!!)^2 \left(1 + \frac{1}{\alpha_k - 1}\right) \times \cdots \times 2 \le (\alpha_k!!)^2 2^{\alpha_k/2}$. For any odd α_k, $\alpha_k! \le (\alpha_k!!)^2 (3/2)^{(\alpha_k - 1)/2} \le (\alpha_k!!)^2 2^{\alpha_k/2}$. As a result, $m_\alpha^\Phi \le \sqrt{2^{\sum_{k=1}^K \alpha_k/2} \prod_{k=1}^K \alpha_k!}$ and $\Phi \in \mathcal{F}^{e+}$.

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

functions generated by finite mixtures of density f_{e+}, where f_{e+} is the density corresponding to the distribution function $F_{e+} \in \mathcal{F}^{e+}$:

$$\mathcal{M}^{f_{e+}} = \left\{ \sum_{i=1}^{m} c_i \frac{1}{\varsigma_i^K} f_{e+} \left(\frac{(\beta_k)_{k=1}^K - \mu_i}{\varsigma_i} \right), \sum_{i=1}^{m} c_i = 1, \ c_i > 0, \ \mu_i \in \mathbb{R}^K, \ \varsigma_i > 0, \ 1 \le i \le m; m \in \mathbb{N} \right\}. \tag{8}$$

COROLLARY 1.

(a) *Suppose that* $G \in \mathcal{M}^{\phi} + \mathcal{M}^{\psi}$ *in (5). Then,* $(\sigma_j(\xi_t; X_t, F, d_i; G))_{j=1}^J$ *is real analytic with respect to* $\xi_t \in \mathbb{R}^J$ *and* $\inf_{\xi_t \in \mathbb{R}^J, F \in \mathcal{F}} r(\xi_t; F, G) > 0$.

(b) *Suppose that* $G \in \mathcal{M}^{\phi} + \mathcal{M}^{\psi}$ *in (5),* $g_j(x_{jt}; \beta_i) = x_{jt}\beta_i$, *and* $F \in \mathcal{F}^e$ *in (6). Then,* $(\sigma_j(\xi_t; X_t, F, d_i; G))_{j=1}^J$ *is real analytic with respect to* $(\xi_t, X_t) \in \mathbb{R}^J \times \mathbb{R}^{J \times K}$ *and* $\inf_{\xi_t \in \mathbb{R}^J, X_t \in \mathbb{R}^{J \times K}} r(\xi_t, X_t; G) > 0$, *where* $r(\xi_t, X_t; G) = \inf_{m \in Supp(F)} r_m(\xi_t, X_t; G)$.

(c) *(Mixed probit) Suppose that* $G \in \mathcal{M}^{\phi}$ *in (4),* $g_j(x_{jt}; \beta_i) = x_{jt}\beta_i$, *and* $F \in \mathcal{M}^{f_{e+}}$ *in (8) with* $F_{e+} \in \mathcal{F}^{e+}$ *in (7). Then,* $(\sigma_j(\xi_t; X_t, F, d_i; G))_{j=1}^J$ *is real analytic with respect to* $(\xi_t, X_t) \in \mathbb{R}^J \times \mathbb{R}^{J \times K}$ *and* $\inf_{\xi_t \in \mathbb{R}^J, X_t \in \mathbb{R}^{J \times K}} r(\xi_t, X_t; G) > 0$.

Corollary 1(a) extends Theorem 1 to any finite mixture of mixed logit and mixed probit models (2). Focusing on the linear index specification, Corollaries 1(b) and 1(c) extend real analyticity with respect to ξ_t to real analyticity with respect to (ξ_t, X_t). Because of the stronger real analytic property of the mixed probit (relative to the mixed logit), Corollary 1(c) extends real analyticity with respect to X_t to any finite mixture of mixed probit models (2) with moments of the random coefficients increasing at a faster rate, such as the squared root of a factorial rate (e.g., the Gaussian distribution or mixtures of it).

An important implication of Corollary 1 is that *any* demand model (2)—not necessarily real analytic—can be approximated, in terms of ξ_t (and X_t), by a sequence of real analytic demand models. This is important because, while the real analytic demand models in Corollary 1 are subject to restrictions (in terms of G and, in Corollaries 1(b) and 1(c), also F and $g_j(x_{jt}; \beta_i)$), finite mixtures of these preserve real analyticity and can approximate any more general demand model (2) that does not need to satisfy the restrictions of Corollary 1. As an example, consider a finite mixture of mixed probit models. According to Nguyen et al. (2020) (their Thm. 5(f)), for any distribution G with density g, we can find a sequence of distributions $G_m^{\phi} \in \mathcal{M}^{\phi}$ such that $\|g - g_m^{\phi}\|_{L_1} \to 0$ as $m \to \infty$. Then,

$$\begin{aligned} \sup_{j, \xi_t} \left| \sigma_j(\xi_t; X_t, F, d_i, G) - \sigma_j(\xi_t; X_t, F, d_i, G_m^{\phi}) \right| &\le \sup_{j, \xi_t} \int \mathbf{1}\{U_{ijt} > U_{ijr}, \forall r \neq j\} dF(\beta_i; d_i) \left| g(\varepsilon) - g_m^{\phi}(\varepsilon) \right| d\varepsilon \\ &\le M_F \|g - g_m^{\phi}\|_{L_1} \\ &\to 0, \end{aligned} \tag{9}$$

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

where $M_F = \max\left\{1, \sup_{\beta\in\mathcal{B}} \frac{dF(\beta;d_i)}{d\beta}\right\}$ and $\mathcal{B}$ is the subset of the support of β_i that excludes any mass point. According to Corollary 1(a), $\sigma_j(\xi_t;X_t,F,d_i,G_m^{\phi})$ is real analytic with respect to $\xi_t \in \mathbb{R}^J$. We can then obtain a uniform approximation of $\sigma_j(\xi_t;X_t,F,d_i,G)$ by a sequence of $\sigma_j(\xi_t;X_t,F,d_i,G_m^{\phi})$ that are real analytic with respect to $\xi_t \in \mathbb{R}^J$. Similarly, we can achieve such uniform approximation jointly in terms of (ξ_t,X_t) on the basis of the real analytic demand models in Corollaries 1(b) and 1(c). The next proposition formalizes these approximation results.

THEOREM 2. *For any G and F with densities g and f defined in $\mathbb{R}^{J+1}$ and $\mathbb{R}^K$, respectively, there exists a sequence of real analytic demand systems $\{\sigma(\xi_t;X_t,F,d_i,G_m)\}_{m=1}^{+\infty}$ and $\{\sigma(\xi_t;X_t,F_m,d_i,G_m)\}_{m=1}^{+\infty}$ with $G_m \in \mathcal{M}^{\phi}+\mathcal{M}^{\psi}$ in (5) and $F_m \in \mathcal{F}^e$ in (6) such that:*

(a) $\lim_{m\to\infty} \sup_{j,\xi_t} \left|\sigma_j(\xi_t;X_t,F,d_i,G)-\sigma_j(\xi_t;X_t,F,d_i,G_m)\right| = 0.$

(b) *Suppose $g_j(x_{jt};\beta_i) = x_{jt}\beta_i$. Then* $\lim_{m\to\infty} \sup_{j,\xi_t,X_t} \left|\sigma_j(\xi_t;X_t,F,d_i,G)-\sigma_j(\xi_t;X_t,F_m,d_i,G_m)\right| = 0.$

(c) *(Mixed probit) Suppose $g_j(x_{jt};\beta_i) = x_{jt}\beta_i$. Then for any G and F with densities g and f defined in $\mathbb{R}^{J+1}$ and $\mathbb{R}^K$, respectively, there exists a sequence of real analytic demand systems $\{\sigma(\xi_t;X_t,F_m,d_i,G_m)\}_{m=1}^{+\infty}$ with $G_m, F_m \in \mathcal{M}^{\phi}$ in (4) such that*

$$\lim_{m\to\infty} \sup_{j,\xi_t,X_t} \left|\sigma_j(\xi_t;X_t,F,d_i,G)-\sigma_j(\xi_t;X_t,F_m,d_i,G_m)\right| = 0.$$

Theorem 2(a) presents a uniform approximation in terms of ξ_t by the real analytic demand models in Corollary 1(a), while Theorems 2(b) and 2(c) restrict attention to linear indices and characterize uniform approximations in terms of (ξ_t,X_t) by the real analytic demand models in Corollaries 1(b) and 1(c), respectively. On the one hand, the approximants in Theorems 2(b) and 2(c) rely on the same linear indices as the demand models to be approximated. On the other hand, the two results rely on distinct families of distributions ($\mathcal{F}^e$ and $\mathcal{M}^{\phi}$, respectively) for the approximation in terms of X_t: G_m and F_m in Theorem 2(b) belong to different families, while they are both Gaussian mixtures in Theorem 2(c).

An immediate implication of the joint approximations in terms of (ξ_t,X_t) in Theorems 2(b) and 2(c) is that *any* random utility model (RUM) in the class considered by McFadden and Train (2000) (see paper for details) can be approximated arbitrarily well by a real analytic demand model.

Remark 4. For any RUM in the class considered by McFadden and Train (2000), there exists a sequence of real analytic demand models as defined in Theorem 2(b) (or 2(c)) that uniformly converges to it in terms of $(\xi_t,X_t) \in \mathbb{R}^J \times \mathbb{R}^{J\times K}$.

Remark 4 builds on McFadden and Train (2000), who show that any RUM can be approximated in terms of (ξ_t,X_t) by a mixed logit model with linear indices (including polynomials of X_t, see Remark 3) and a flexible F. Note that the mixed logit approximant proposed by McFadden and Train (2000) is more flexible in terms of F than the real analytic demand models considered in Corollaries 1(b) and 1(c), and thus its real analyticity does not follow from these. However, using Theorem 2(b) (or 2(c)) and the fact that polynomials of X_t are also real analytic, one can approximate uniformly and arbitrarily well any such flexible mixed logit approximant by a real analytic demand model with the same index structure, therefore uniformly approximating the original RUM (see Appendix C).

To summarize, Theorem 2 and Remark 4 imply that, when dealing with demand models generated by (2) or any RUM in the class considered by McFadden and Train (2000), one can restrict attention to a subset of real analytic demand models without any loss of precision. In the reminder of the paper, we discuss the econometric advantages of dealing with such real analytic demand models in terms of both identification and numerical implementation. To facilitate exposition, we suppress from our notation d_i when referring to σ_j and the subscript i when referring to p_{ijt}.

4. IMPLICATIONS FOR IDENTIFICATION

In this section, we discuss how real analyticity facilitates the nonparametric and semi-nonparametric identification of demand model (2) with linear indices $g_j(x_{jt};\beta_i) = x_{jt}\beta_i$. In this case, (2) can be equivalently expressed as $\sigma_j(\delta_t;X_t,F,G)$, a function of the J-dimensional vector of average utilities $\delta_t = (\delta_{jt})_{j=1}^J$ with jth element $\delta_{jt} = x_{jt}\beta + \xi_{jt}$, where β is the population average of the vector of random coefficient β_i and $\beta_i = \beta + \Delta\beta_i$. Define $\sigma(\delta_t;X_t,F,G) = (\sigma_j(\delta_t;X_t,F,G))_{j=1}^J$. Analogously, the real analyticity of $\sigma(\delta_t;X_t,F,G)$ with respect to ξ_t (and X_t) can be equivalently restated with respect to δ_t (and X_t).

4.1. Nonparametric Identification

Researchers are often interested in identifying $\sigma(\delta_t;X_t,F,G)$ as a function of (δ_t,X_t), what can be referred to as the nonparametric approach (Berry and Haile, 2014, 2018). In this approach, one wishes to remain agnostic about the functional forms of G and F, so to reduce the potential for misspecification. This flexibility allows demand model (2) to subsume a broad range of consumer preferences and behaviors, including complex substitution patterns, continuous quantity choices, complementarities across products, or consumer inattention, and consequently it allows researchers to perform more realistic counterfactual simulations, for example, hypothetical mergers or the introduction of new taxes (Compiani, 2022).

The real analyticity of demand model (2) implies a powerful form of *identification extensibility*, which is crucial to enable researchers to perform hypothetical counterfactuals on the basis of the nonparametric approach. Without further restrictions on $\sigma(\delta_t;X_t,F,G)$, the nonparametric identification of $\sigma(\delta_t;X_t,F,G)$ in the support of (δ_t,X_t) spanned by the data, Ω, only allows the researcher to predict counterfactual outcomes within the same support. However, it is possible for some counterfactual outcomes to fall outside of Ω, for example, new equilibrium prices that significantly increase after a hypothetical merger or the introduction of a new tax. Identifying such counterfactual outcomes is essentially an extrapolation exercise for which the nonparametric identification of $\sigma(\delta_t; X_t, F, G)$ in Ω may not be sufficient in the absence of additional restrictions on $\sigma(\delta_t; X_t, F, G)$.[7] The identification extensibility due to real analyticity proved in the next proposition overcomes this challenge: the nonparametric identification of a real analytic $\sigma(\delta_t; X_t, F, G)$ is sufficient for any extrapolation outside of Ω.

COROLLARY 2 (Identification extensibility). *Denote by Ω the support of (δ_t, X_t). Suppose that $G \in \mathcal{M}^{\phi} + \mathcal{M}^{\psi}$ in (5) and that Ω contains an open subset.*

(a) *Denote by Ω_δ the domain of δ_t. Then, for any X_t,*

$$\sigma(\delta_t; X_t, F, G) \text{ is identified for } \delta_t \in \Omega_\delta \implies \sigma(\delta_t; X_t, F, G) \text{ is identified for } \delta_t \in \mathbb{R}^J.$$

(b) *Suppose $F \in \mathcal{F}^e$ in (6). Then,*

$$\sigma(\delta_t; X_t, F, G) \text{ is identified for } (\delta_t, X_t) \in \Omega$$
$$\implies \sigma(\delta_t; X_t, F, G) \text{ is identified for } (\delta_t, X_t) \in \mathbb{R}^J \times \mathbb{R}^{J\times K}.$$

(c) *Suppose $G, F \in \mathcal{M}^{\phi}$ in (4). Then,*

$$\sigma(\delta_t; X_t, F, G) \text{ is identified for } (\delta_t, X_t) \in \Omega$$
$$\implies \sigma(\delta_t; X_t, F, G) \text{ is identified for } (\delta_t, X_t) \in \mathbb{R}^J \times \mathbb{R}^{J\times K}.$$

Corollary 2 holds due to two properties highlighted in Theorem 1 and Corollary 1. First, because $\sigma(\delta_t; X_t, F, G)$ is real analytic for any $(\delta_t, X_t) \in \mathbb{R}^J \times \mathbb{R}^{J\times K}$, then for any $\|(\delta_t', X_t') - (\delta_t, X_t)\| < r(\delta_t, X_t; F, G)$, where $r(\delta_t, X_t; F, G)$ is the radius of convergence at (δ_t, X_t), we have

$$\sigma(\delta_t'; X_t', F, G) = \sigma(\delta_t; X_t, F, G) + \sum_{L=1}^{\infty} \sum_{\sum_{k=1}^{J(P+1)} \alpha_k = L} \frac{\prod_{j=1}^{J} (\delta_{jt}' - \delta_{jt})^{\alpha_j} \prod_{j\le J; p\le P} (x_{jp't} - x_{jpt})^{\alpha_{pJ+j}}}{\prod_{k=1}^{J(P+1)} \alpha_k!} \times \prod_{j=1}^{J} \partial_{\delta_j}^{\alpha_j} \prod_{j\le J; p\le P} \partial_{x_{jp}}^{\alpha_{pJ+j}} \sigma(\delta_t; X_t, F, G), \tag{10}$$

or, in words, the Taylor series of $\sigma(\delta_t'; X_t', F, G)$ at (δ_t, X_t) converges to $\sigma(\delta_t'; X_t', F, G)$. Because $\sigma(\delta_t; X_t, F, G)$ is identified for $(\delta_t, X_t) \in \Omega$, we then identify all its derivatives $\prod_{j=1}^{J} \partial_{\delta_j}^{\alpha_j} \prod_{j\le J; p\le P} \partial_{x_{jp}}^{\alpha_{pJ+j}} \sigma(\delta_t; X_t, F, G)$ for $(\delta_t, X_t) \in \Omega$. Second, we know from Theorem 1 and Corollary 1 that $r = \inf_{\delta_t \in \mathbb{R}^J, X_t \in \mathbb{R}^{J\times K}} r(\delta_t, X_t; F) > 0$. As a result, when (δ_t, X_t) approaches the boundary of Ω, we can

[7] Differently, the semi-nonparametric approach we discuss below in Section 4.2 does not suffer from this extrapolation problem: once F is identified, one can evaluate demand model (2) at any value of (δ_t, X_t).

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

extrapolate $\sigma(\delta_t';X_t',F,G)$ using the identified right-hand side of (10) for $\|(\delta_t',X_t')-(\delta_t,X_t)\|<r$ and $(\delta_t',X_t')\notin\Omega$, identifying $\sigma(\delta_t';X_t',F,G)$ in $\Omega_r^{(1)}=\{(\delta_t';X_t'):\ \exists(\delta_t;X_t)\in\Omega \text{ such that } \|(\delta_t';X_t')-(\delta_t;X_t)\|<r\}$. By repeating the procedure, we identify $\sigma(\delta_t;X_t,F,G)$ in $\mathbb{R}^J\times\mathbb{R}^{J\times K}$.[8,9]

To illustrate the usefulness of Corollary 2 for the nonparametric identification of hypothetical counterfactuals, we discuss two common examples from industrial organization (Nevo, 2000; Dubois et al., 2020) that are typically hard to evaluate absent identification extensibility.

Example 1 (Merger). Suppose $J=2$, with each product being sold by a different firm who chooses its price $p_j, j=1,2$. Suppose the two products are substitutes and the cross-price elasticities are always positive. The variable profits of the two firms are:

$$\pi_1(p_1;p_2)=(p_1-c_1)\sigma_1(p_1;p_2),$$
$$\pi_2(p_2;p_1)=(p_2-c_2)\sigma_2(p_2;p_1).$$

The observed prices p_1^* and p_2^* are generated by a simultaneous Bertrand pricing game with constant marginal costs $c_1<p_1^*$ and $c_2<p_2^*$, respectively. Then, the corresponding first-order conditions (FOCs) are

$$\frac{\partial\pi_1(p_1^*;p_2^*)}{\partial p_1}=0 \iff \frac{\partial\sigma_1(p_1^*,p_2^*)}{\partial p_1}(p_1^*-c_1)+\sigma_1(p_1^*,p_2^*)=0,$$
$$\frac{\partial\pi_2(p_2^*;p_1^*)}{\partial p_2}=0 \iff \frac{\partial\sigma_2(p_1^*,p_2^*)}{\partial p_2}(p_2^*-c_2)+\sigma_2(p_1^*,p_2^*)=0.$$

In the case of a hypothetical merger between the two firms, the merged entity maximizes the joint profits generated by both products, $\pi_1(p_1;p_2)+\pi_2(p_2;p_1)$, and post-merger equilibrium prices (p_1^m,p_2^m) satisfy the following FOCs:

$$\begin{aligned}&\frac{\partial\sigma_1(p_1,p_2)}{\partial p_1}(p_1-c_1)+\frac{\partial\sigma_2(p_1,p_2)}{\partial p_1}(p_2-c_2)+\sigma_1(p_1,p_2)=0,\\&\frac{\partial\sigma_2(p_1,p_2)}{\partial p_2}(p_2-c_2)+\frac{\partial\sigma_1(p_1,p_2)}{\partial p_2}(p_1-c_1)+\sigma_2(p_1,p_2)=0.\end{aligned} \qquad \textbf{(11)}$$

Because $\frac{\partial\sigma_1}{\partial p_2}>0$ and $\frac{\partial\sigma_2}{\partial p_1}>0$, both FOCs in (11) are then positive when evaluated at the observed prices (p_1^*,p_2^*):

[8]In the same spirit but in the context of identification, Allen and Rehbeck (2020) obtain the nonparametric identification of indirect utility functions by relying on the unique continuation property implied by real analyticity (see their Cor. 3). Similarly, one of the key identifying assumptions in Fox and Gandhi (2016), Assumption 4 (p. 127), is satisfied by multivariate real analytic utility functions again because of the unique continuation property (see their appendix).

[9]This extension procedure also implies a constructive estimation of $\sigma(\delta_t;X_t,F,G)$ for $(\delta_t,X_t)\notin\Omega$ by estimating its (higher-order) derivatives and extrapolating using its estimated Taylor series (up to a finite order). This is beyond the scope of the paper and we leave it for future research.

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

$$\frac{\partial\sigma_1(p_1^*,p_2^*)}{\partial p_1}(p_1^*-c_1)+\frac{\partial\sigma_2(p_1^*,p_2^*)}{\partial p_1}(p_2^*-c_2)+\sigma_1(p_1^*,p_2^*)=\frac{\partial\sigma_2(p_1^*,p_2^*)}{\partial p_1}(p_2^*-c_2)>0,$$
$$\frac{\partial\sigma_2(p_1^*,p_2^*)}{\partial p_2}(p_2^*-c_2)+\frac{\partial\sigma_1(p_1^*,p_2^*)}{\partial p_2}(p_1^*-c_1)+\sigma_2(p_1^*,p_2^*)=\frac{\partial\sigma_1(p_1^*,p_2^*)}{\partial p_2}(p_1^*-c_1)>0.$$

Under standard regularity conditions on $(\sigma_j)_{j=1}^J$ (e.g., single-peak property), this implies that the merged entity has an incentive to increase post-merger prices relative to (p_1^*,p_2^*) and therefore $p_j^m>p_j^*$ for $j=1,2$. If products 1 and 2 are strong substitutes (i.e., $\frac{\partial\sigma_2(p_1^*,p_2^*)}{\partial p_1}$ and $\frac{\partial\sigma_1(p_2^*;p_1^*)}{\partial p_2}$ are very positive), then p_j^m may exceed the maximal value of p_j^* observed in the data and the corresponding counterfactual demand $(\sigma_1(p_1^m;p_2^m),\sigma_2(p_2^m;p_1^m))$ may not be identified.

Example 2 (Tax). Consider a similar setting to that in Example 1 but with only one product (so we suppress the subscript referring to products). In the observed data, the firm chooses the price by maximizing variable profit:

$$\pi_1(p)=(p-c)\sigma(p).$$

Then the observed optimal price p^* satisfies the FOC:

$$\frac{\partial\sigma(p^*)}{\partial p}(p^*-c)+\sigma(p^*)=0.$$

Suppose that the government announces an ex-factory tax $\tau>0$ such that the final price consumers face will be $p_\tau=p+\tau$. Then, the equilibrium price p_τ^* satisfies

$$\frac{\partial\sigma(p)}{\partial p}(p-c-\tau)+\sigma(p)=0, \tag{12}$$

where the ex-factory tax τ is equivalent to an increase in the marginal cost of production. Because $\frac{\partial\sigma(p)}{\partial p}<0$ (i.e., the law of demand), then FOC (12) evaluated at $p=p^*$ (the observed optimal price in the absence of the tax) is positive:

$$\frac{\partial\sigma(p^*)}{\partial p}(p^*-c-\tau)+\sigma(p^*)=-\frac{\partial\sigma(p^*)}{\partial p}\tau>0.$$

As a result, if the tax τ is large enough, p_τ^* may exceed the maximal value of p^* observed in the data and the corresponding $\sigma(p_\tau^*)$ may not be identified.

4.2. Semi-Nonparametric Identification

Real analyticity also facilitates the semi-nonparametric identification of demand model (2) with linear indices $g_j(x_{jt};\beta_i)=x_{jt}\beta_i$. In this approach, the researcher takes G as known (e.g., i.i.d. Gumbel or Gaussian) and aims at identifying the distribution of the random coefficients F. The main advantage of this approach is that knowledge of G and of F allows the quantification of the distribution of welfare effects resulting from hypothetical counterfactuals, a task for which the nonparametric approach may be inadequate (Compiani, 2022).

Without real analyticity, the semi-nonparametric identification of discrete choice model (2) with linear indices usually requires large support for X_t (restrictive for example with price variables, known to be positive) and prevents the inclusion of interactions among covariates (Fox et al., 2012; Fox and Gandhi, 2016; Masten, 2018). These standard assumptions, in some cases, limit the economic content of discrete choice models. For example, the approximation result by McFadden and Train (2000) mentioned after Remark 4, crucially relies on the inclusion of interactions among covariates (see p. 466). The identification results by Fox et al. (2012), however, explicitly rule out interactions among covariates (Assumptions 2 and 3, pp. 207–208).

Differently, as shown by Wang (2023), when G is i.i.d. Gumbel, the identification of F can be obtained by relying on at most one single variation in X_t (i.e., the support of X_t is not singleton).[10] This support requirement on X_t substantially relaxes standard conditions routinely used in the literature and is satisfied in most settings. In particular, it allows for interactions among covariates as required by the approximation result by McFadden and Train (2000). Wang's (2023) identification result for mixed logit models crucially relies on the identification of $\sigma(\delta_t; X_t, F, G)$ as a function of $\delta_t \in \mathbb{R}^J$. According to the identification extensibility property from Corollary 2(a), when $\sigma(\delta_t; X_t, F, G)$ is real analytic, this can be achieved as long as $\sigma(\delta_t; X_t, F, G)$ is identified for δ_t in a (bounded) open set. Importantly, this local condition is not only weaker, but also consistent with standard economic models. For instance, Wang (2023) shows that, in the context of Berry et al. (1995), it is implied by a simultaneous price-setting game of complete information among producers.

The next proposition extends Wang's (2023) identification result for mixed logit models with linear indices to any demand model (2) with linear indices and $G \in \mathcal{M}^\phi + \mathcal{M}^\psi$.

COROLLARY 3. *Suppose that* $G \in \mathcal{M}^\phi + \mathcal{M}^\psi$ *in (5) and the following two conditions hold:*

1. $\sigma(\delta_t; X_t, F, G)$ *is identified as a function of* $(\delta_t, X_t) \in \Omega$, *where the domain of* δ_t, Ω_δ, *contains an open set in* $\mathbb{R}^J$.
2. *There exists* $(\delta_t, X_t) \in \Omega$ *such that* X_t *is of full column rank.*

Then, F is identified.

5. IMPLICATIONS FOR NUMERICAL IMPLEMENTATION

Depending on the type of data, different methods can be used to estimate model (2). We discuss the implications of real analyticity for the numerical implementation of two methods routinely used in the empirical literature for the estimation of demand models: on the one hand, the parametric BLP (Berry et al., 1995) and

[10] For details, see Wang's (2023) Theorems 1 and 2 and the discussion therein.

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

the semi-nonparametric approach (Wang, 2023) dealing with aggregate market-level data; on the other, the MLE dealing with disaggregate individual-level data (Goolsbee and Petrin, 2004; Train and Winston, 2007; Dubois et al., 2020). In both cases, we focus on the linear index specification $g_j(x_{jt};\beta_i) = x_{jt}\beta_i$.

5.1. BLP and Semi-Nonparametric Implementation

In the context of demand estimation with aggregate market-level data, both the parametric BLP and the semi-nonparametric approach require the computation of a demand inverse for each market and at each iteration of the GMM minimization. For given market t, observed p_t, and guess F', one must look for a δ_t' such that[11]

$$p_t = \sigma(\delta_t'; X_t, F', G). \tag{13}$$

This can be numerically challenging on a computer (Knittel and Metaxoglou, 2014), motivating different numerical approaches in the economics literature, such as Fixed Point (FP) approaches (Berry et al., 1995), MPEC (Su and Judd, 2012; Dubé et al., 2012a), and the use of an approximate inverse (Salanié and Wolak, 2019).

The real analyticity of $\sigma(\delta_t; X_t, F', G)$ with respect to δ_t mitigates this challenge by guaranteeing the desirable numerical performance of Newton-Raphson (NR) methods. In the NR implementation of the unique solution of (13), starting from $\delta^{(0)}$, one performs the following iteration from $n = 1$ until numerical convergence:

$$\delta^{(n)} = \delta^{(n-1)} - \left[\partial_\delta \sigma(\delta^{(n)}; X_t, F', G)\right]^{-1} \left(\sigma(\delta^{(n)}; X_t, F', G) - p_t\right), \tag{14}$$

where ∂_δ refers to the derivative with respect to δ. The common wisdom regarding (14) is that if $\delta^{(0)}$ is "close" to the solution δ_t' (i.e., $\delta^{(0)}$ is in the basin of attraction of δ_t'), then (14) can achieve quadratic convergence as long as $\sigma(\delta; X_t, F', G)$ is twice continuously differentiable with respect to δ.[12] Achieving quadratic convergence then crucially relies on selecting a starting value $\delta^{(0)}$ which is close to δ_t', but the extent of such proximity typically depends on knowledge of the true δ_t' (or $\partial_\delta \sigma(\delta_t'; X_t, F', G)$) that is not available to the researcher before estimation. In addition, numerical convergence is usually determined by the researcher, who sets some small tolerance for the step lengths $\left\|\delta^{(n+1)} - \delta^{(n)}\right\|$ and $\left\|\sigma(\delta^{(n)}; X_t, F', G) - p_t\right\|$. However, there is little theoretical guidance on how to choose such levels of tolerance, which are often calibrated using heuristic rules of thumb. In fact, despite achieving numerical convergence on the basis of such rules of thumb, it is still possible for (14) to be divergent. Together, unguided choices

[11] When demographics d_i enter σ, demand system (13) is defined as

$$p_t = \int \sigma(\delta_t; X_t, F, d_i, G) d\Pi(d_i).$$

The result presented in this section solely depends on the real analyticity with respect to δ_t, that is, Corollary 1(a). Train and Winston (2007) also implement inverses of this demand system in the context of MLE with individual-level data. Instead of inverting market shares, they invert choice probabilities obtained from observed individual choices.

[12] Quadratic convergence means that there exists a constant M such that $\|\delta^{(n+1)} - \delta^{(n)}\| \leq M\|\delta^{(n)} - \delta^{(n-1)}\|^2$.

of starting values and of stopping criteria represent a challenge for the numerical convergence of NR methods to the unique solution of (13) (Dubé, Fox, and Su, 2012b; Lee and Seo, 2016; Conlon and Gortmaker, 2020).

We show that the real analyticity of $\sigma(\delta; X_t, F', G)$ gives rise to simple and verifiable sufficient conditions that do not require knowledge of the true δ_t' (or $\partial_\delta \sigma(\delta_t'; X_t, F', G)$) and that guarantee the convergence of (14) to the unique solution of (13). Following Smale (1986), we define $\delta^{(0)}$ as an *approximate zero* of (13) if $\delta^{(0)}$ satisfies[13]

$$\left\| \delta^{(n)} - \delta^{(n-1)} \right\| \leq \left(\frac{1}{2} \right)^{2^{n-1}-1} \left\| \delta^{(1)} - \delta^{(0)} \right\|, \text{ for } n \geq 1. \tag{15}$$

If the iterations in (14) start with an approximate zero $\delta^{(0)}$, then $\delta^{(n)}$ is guaranteed to converge to δ_t' at supra-exponential rate, $\left\| \delta^{(n)} - \delta_t' \right\| \leq \frac{7}{4} \left(\frac{1}{2} \right)^{2^{n-1}} \left\| \delta^{(1)} - \delta^{(0)} \right\|$ (Prop. 1 at p. 188 of Smale, 1986), a rate of the same order as the rate of quadratic convergence (see Rheinboldt, 1988 for details) and that does not depend on the dimension of the demand system (i.e., the number of alternatives J). For example, with $n = 6$, we have $\frac{7}{4} \left(\frac{1}{2} \right)^{2^{n-1}} \approx 4.07 \times 10^{-10}$.

When $\sigma(\delta_t; X_t, F', G)$ is real analytic with respect to $\delta_t \in \mathbb{R}^J$, Smale's (1986) Theorem A characterizes a sufficient condition for $\delta^{(0)}$ to be an approximate zero:

$$\alpha(\delta^{(0)}) = \left\| \left[\partial_\delta \sigma(\delta^{(0)}; X_t, F', G) \right]^{-1} (\sigma(\delta^{(0)}; X_t, F', G) - p_t) \right\| \gamma(\delta^{(0)}) < \alpha_0 \approx 0.130707, \tag{16}$$

where

$$\gamma(\delta) := \sup_{k>1} \left\| \left[\partial_\delta \sigma(\delta; X_t, F', G) \right]^{-1} \frac{\partial_\delta^k \sigma(\delta; X_t, F', G)}{k!} \right\|^{\frac{1}{k-1}},$$

with the operator ∂_δ^k denoting the kth derivative with respect to δ as a k-linear map (see Chap. XIII of Lang, 2012 for a definition), and the norm $\|\cdot\|$ of a linear operator $L : E \to W$ being defined as

$$\|L\| := \sup_{v \in E, \|v\|_E = 1} \|L(v)\|_W,$$

with $\|\cdot\|_E$ and $\|\cdot\|_W$ denoting norms defined in the spaces E and W, respectively. In theory, one can compute each component of $\alpha(\delta^{(0)})$ and then compare it to α_0 to verify whether $\delta^{(0)}$ is an approximate zero. In practice, while computing $\left[\partial_\delta \sigma(\delta; X_t, F', G) \right]^{-1} \sigma(\delta; X_t, F', G)$ is straightforward, computing $\gamma(\delta)$ is more involved due to the higher-order derivatives and the sup operator. However, relying on our previous results (Lemmas A2 and A3 in Appendix A), when $G \in \mathcal{M}^\phi + \mathcal{M}^\psi$, we can derive simple upper bounds for the higher-order

[13] For the original definition of an approximate zero, see page 187 of Smale (1986). A value of δ that is an approximate zero is also in the basin of attraction of the unique solution of (13), but the converse does not need to be true.

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

derivatives of $\sigma(\cdot;X_t,F',G)$ and for $\gamma(\delta)$. This allows us to obtain a practically simpler (but stronger than (16)) sufficient condition for approximate zeros that circumvents the computational complexity in the general formulation of Theorem A by Smale (1986). Denote $\underline{\varsigma} := \min\{\varsigma_{11},\ldots,\varsigma_{m_1 1},\varsigma_{12},\ldots,\varsigma_{m_2 2}\}$ as the minimal standard variance among the distributions in the mixture that defines $G \in \mathcal{M}^{\phi}+\mathcal{M}^{\psi}$ (see (B.1)).

COROLLARY 4. *Suppose that $G \in \mathcal{M}^{\phi}+\mathcal{M}^{\psi}$ in (5). Then, for $\delta \in \mathbb{R}^J$,*

$$\gamma(\delta) \leq \frac{J\|[\partial_\delta \sigma(\delta;X_t,F',G)]^{-1}\|\sup_{k>1}\left[\left(r\left(\frac{C}{\sqrt{2\pi}}\right)^J \frac{1}{\sqrt{k!}}+(1-r)e^{1/(e-1)}(e-1)^k\right)\right]^{\frac{1}{k-1}}}{2\min\{1,\underline{\varsigma}^2\}},$$

where C is a constant defined in (A.11) in Appendix A (that does not depend on J) and r is the probability of the Gaussian mixtures in G (see (B.1) in Appendix B). Moreover, any δ is an approximate zero as long as

$$\begin{aligned}\bar{\alpha}(\delta) := &\left\|\left[\partial_\delta \sigma(\delta;X_t,F',G)\right]^{-1}(\sigma(\delta;X_t,F',G)-p_t)\right\| \\ &\times \frac{J\|[\partial_\delta \sigma(\delta;X_t,F',G)]^{-1}\|\sup_{k>1}\left[\left(r\left(\frac{C}{\sqrt{2\pi}}\right)^J \frac{1}{\sqrt{k!}}+(1-r)e^{1/(e-1)}(e-1)^k\right)\right]^{\frac{1}{k-1}}}{2\min\{1,\underline{\varsigma}^2\}} \\ &< \alpha_0.\end{aligned}$$

Given δ, one can easily obtain $\bar{\alpha}(\delta)$ by computing $\left\|\left[\partial_\delta \sigma(\delta;X_t,F',G)\right]^{-1}\sigma(\delta;X_t,F',G)\right\|$ and $\left\|\left[\partial_\delta \sigma(\delta;X_t,F',G)\right]^{-1}\right\|$, where $[\partial_\delta \sigma(\delta;X_t,F',G)]^{-1}$ is the inverse of the Jacobian matrix $\partial_\delta \sigma(\delta;X_t,F',G)$ and $\|[\partial_\delta \sigma(\delta;X_t,F',G)]^{-1}\|$ is its maximal eigenvalue. Then, it is straightforward to establish whether the starting value $\delta^{(0)}$ is an approximate zero by verifying whether $\bar{\alpha}(\delta^{(0)}) < \alpha_0$, an operation that does not require any NR iteration.

More in general, at any iteration n of (14) one can compute

$$\begin{aligned}\bar{\alpha}(\delta^{(n)}) = &\left\|\delta^{(n+1)}-\delta^{(n)}\right\| \\ &\times \left\|[\partial_\delta \sigma(\delta^{(n)};X_t,F',G)]^{-1}\right\| \frac{\sup_{k>1}\left[\left(r\left(\frac{C}{\sqrt{2\pi}}\right)^J \frac{1}{\sqrt{k!}}+(1-r)e^{1/(e-1)}(e-1)^k\right)\right]^{\frac{1}{k-1}}}{2\min\{1,\underline{\varsigma}^2\}} J\end{aligned}$$

and verify whether the step length

$$\begin{aligned}&\left\|\delta^{(n+1)}-\delta^{(n)}\right\| < \\ &\Delta^* := \frac{2\alpha_0 \min\{1,\underline{\varsigma}^2\}}{J\left\|[\partial_\delta \sigma(\delta^{(n)};X_t,F',G)]^{-1}\right\|\sup_{k>1}\left[\left(r\left(\frac{C}{\sqrt{2\pi}}\right)^J \frac{1}{\sqrt{k!}}+(1-r)e^{1/(e-1)}(e-1)^k\right)\right]^{\frac{1}{k-1}}}.\end{aligned}$$

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

When this condition holds, one can conclude that $\delta^{(n)}$ is an approximate zero and so that supra-exponential convergence is guaranteed thereafter, ruling out any improper numerical convergence (i.e., $\|\delta^{(n+1)}-\delta^{(n)}\|$ is "small" while (14) is divergent). The computation of Δ^* is usually immediate. For example, in the case of the mixed logit model it can be simply approximated by (see Appendix E.1 for details)

$$\Delta^* \approx \frac{\alpha_0 p_{0t} \min_{1\leq j\leq J}\{p_{jt}\}}{2.64J}, \tag{17}$$

where $p_{0t} = 1-\sum_{j=1}^{J} p_{jt}$. For instance, with $J = 10$, $\min_{1\leq j\leq J} p_{jt} = 0.05$, and $p_{0t} = 0.4$, we would obtain $\Delta^* \approx 10^{-4}$.

In practice, while NR algorithm (14) converges to the unique solution much faster than classic FP algorithms from a starting value that is an approximate zero (due to the supra-exponential convergence of (14) versus the exponential convergence of FP algorithms based on contraction mappings), in general it is not guaranteed that *any* $\delta^{(0)}$ is an approximate zero or even in the basin of attraction of the unique solution (see footnote 13). As is well known, when $\delta^{(0)}$ is not in the basin of attraction of the unique solution, then (14) may not converge. In contrast, the convergence of classic FP algorithms based on contraction mappings is guaranteed from *any* starting value, even though it may take a long time. Some researchers address this trade-off by proposing Hybrid Algorithms (HAs) that combine the relative advantages of both FP and NR algorithms: the numerical search starts with a FP algorithm based on a contraction mapping to generate a starting value in the basin of attraction of the unique solution and then switches to a NR algorithm until convergence.[14] In the context of demand inverses, such HA can be expressed as the following.

Algorithm 1 (Hybrid Algorithm). Starting at $\delta^{(0)}$,

FP step. From $n = 0$, update $\delta^{(n)}$ to $\delta^{(n+1)}$ using a FP algorithm based on a contraction mapping until $n = N-1$.

NR step. From $n = N$, update $\delta^{(n)}$ to $\delta^{(n+1)}$ using a NR (14) until numerical convergence.

The number of iterations in the FP step, N, is calibrated by the practitioner. If N is "too" small, the starting value for the NR step, $\delta^{(N+1)}$, may not be in the basin of attraction of the unique solution and therefore the convergence of $(\delta^{(n)})_{n\geq N+1}$ may not be guaranteed. Differently, if N is "too" large, $\delta^{(N+1)}$ is likely to be in the basin of attraction and $(\delta^{(n)})_{n\geq N+1}$ to converge, but the total execution time of Algorithm 1 could be "too" long due to unnecessary extra iterations in the FP step instead of an earlier switch to the NR step.

Corollary 4 provides theoretical guidance for the calibration of N in the context of Algorithm 1. Because of the supra-exponential rate of convergence of (14) after

[14] See Rust (1987) and Iskhakov et al. (2016) for examples in the setting of dynamic discrete choice models.

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

reaching an approximate zero, we propose an implementation of Algorithm 1 that switches from the FP step to the NR step when $\delta^{(N)}$ is guaranteed to be an approximate zero. In the case of the mixed logit model, Corollary 4 implies the following sufficient condition for $\delta^{(N)}$ in the FP step to be an approximate zero (see Appendix E.1 for details):

$$\left\|\delta^{(N+1)} - \delta^{(N)}\right\| < \Delta_{\mathrm{H}}^* = \Delta^* p_{0t}, \tag{18}$$

where $p_{0t} = 1 - \sum_{j=1}^{J} p_{jt}$. Because an approximate zero is guaranteed to be in the basin of attraction of the unique solution, but not the converse (see footnote 13), it is also plausible for more lenient criteria than (18) to work well in practice. As we illustrate in Section 6, we find in Monte Carlo simulations that Algorithm 1 with N calibrated according to our proposed criteria always converges from "distant" starting values (i.e., starting values from which the NR algorithm alone does not converge) and it is faster than classic FP implementations of demand inverses commonly used in practice.

5.2. MLE Implementation

In the context of demand estimation with disaggregate individual-level data, researchers often rely on (parametric) MLE by assuming that G is known and specifying $\beta_i = \Sigma v_i$, where the distribution of $v_i \in \mathbb{R}^{P\times 1}$, F_v, is given (e.g., multivariate Gaussian) and Σ is an unknown matrix of size $K \times P$. Consider a cross-sectional setting in which for each cell $t = 1, \ldots, T$, the researcher observes I individuals each making one choice, denoted by $y_{it} \in \{0, 1, \ldots, J\}$, alternatives' characteristics $(x_{jt})_{j=1}^{J}$, and demographic information d_i. Given these data, the researcher wishes to estimate the "fixed effects" $\xi = (\xi_{jt})_{1\le j\le J, 1\le t\le T}$ and the parameters Σ of the distribution of random coefficients F_v.[15,16] Denote by $\ell_{it}(\xi, \Sigma)$ the log-likelihood function for individual i in cell t:[17]

$$\ell_{it}(\xi, \Sigma) := \ell(\xi, \Sigma; x_{it}, y_{it}) = \ln \int \prod_{j=1}^{J} \left[\sigma_{ij}(\xi_t + x_{it}\Sigma v_i; G)\right]^{\mathbf{1}\{y_{it}=j\}} dF_v(v_i),$$

where $\sigma_{ij}(\cdot; G)$ is the choice probability function of an individual with random coefficients $\beta_i = \Sigma v_i$, and $\mathscr{L}(\xi, \Sigma) = \sum_{i,t} \ell_{it}(\xi, \Sigma)$. In the NR implementation of

[15] The population mean of $x_{jt}\beta_i$ is absorbed by ξ_{jt} and is not a parameter to be estimated. The results in this section also hold if one observes x_{ijt} rather than x_{jt}. Moreover, it is possible that the researcher has panel data and observes multiple choices for the same individual over time. This complicates the individual likelihood function (see footnote 17) but the results in this section continue to apply.

[16] When T is large enough (i.e., a nonlinear panel model with large I and large T), one can also directly estimate $(\beta_i)_{i=1}^{I}$ rather than the distribution F. See Dubois et al. (2020) for such an empirical specification and Mugnier and Wang (2022) for theoretical results. In this case, model (2) is often simpler and real analytic (e.g., simple multinomial logit or probit) and the results in this section apply.

[17] In a panel setting, instead of having an it-specific log-likelihood function, we have an i-specific log-likelihood because β_i is common among the choices over time of the same individual:

$$\ell_i(\xi, \Sigma) = \ln \int \prod_{t=1}^{T} \prod_{j=1}^{J} \left[\sigma_{ij}(\xi_t + x_{it}\Sigma v_i; G)\right]^{\mathbf{1}\{y_{it}=j\}} dF_v(v_i).$$

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

this likelihood maximization, starting from $(\xi^{(0)}, \Sigma^{(0)})$ one performs the following iteration from $n = 1$ until numerical convergence:

$$(\xi^{(n+1)}, \Sigma^{(n+1)}) = (\xi^{(n)}, \Sigma^{(n)}) - \left[\partial^2_{(\xi,\Sigma)}\mathscr{L}(\xi^{(n)}, \Sigma^{(n)})\right]^{-1} \partial_{(\xi,\Sigma)}\mathscr{L}(\xi^{(n)}, \Sigma^{(n)}). \quad \textbf{(19)}$$

Similar to (14), under standard regularity conditions (such as the non-singularity of $\partial^2_{(\xi,\Sigma)}\mathscr{L}(\xi, \Sigma)$ at a true solution), NR method (19) can achieve quadratic convergence by starting from a $(\xi^{(0)}, \Sigma^{(0)})$ that is "close" to a true solution as long as ℓ_{it} (and therefore $\mathscr{L}$) is "smooth enough." However, the practical performance of NR method (19) is subject to the same challenges mentioned in the previous section in the context of (14): the appropriate choices of a starting point and of a stopping criterion usually depend on prior knowledge of the demand model one is trying to estimate. For reasons similar to those behind Corollary 4, the real analyticity of ℓ_{it} with respect to (ξ, Σ) alleviates both such challenges also in the implementation of MLE. Denote $a := \inf_{1 \le i \le I, 1 \le t \le T, 1 \le j \le J} \ell(\xi, \Sigma; x_{it}, j)$ and $\bar{x} = \max_{i,t} |x_{it}|_{\max}$, where $|\cdot|_{\max}$ indicates the maximal absolute value of the elements in x_{it}.

COROLLARY 5. *Suppose that $G \in \mathcal{M}^\phi + \mathcal{M}^\psi$ in (5) and that $F_v \in \mathcal{F}^e$ in (6), or alternatively that $G, F_v \in \mathcal{M}^\phi$ in (4). Then, for (ξ, Σ),*

$$\gamma(\xi,\Sigma) \le \frac{1}{IT}\left\|\left[\frac{\partial^2_{(\xi,\Sigma)}\mathscr{L}(\xi,\Sigma)}{IT}\right]^{-1}\right\| \max\left\{1, \left\|\left[\frac{\partial^2_{(\xi,\Sigma)}\mathscr{L}(\xi,\Sigma)}{IT}\right]^{-1}\right\|\right\}$$
$$\times \left(\max\left\{1, \frac{A_v J(e-1)\underline{\varsigma}^{-1}\bar{x}(a+1)^{1/(JT+KP)}}{(a+1)^{1/(JT+KP)}-1}\right\}\right)^3 \left(\frac{JT+KP}{3}\right)^{3/2},$$

where A_v is the constant corresponding to F_v when $F_v \in \mathcal{F}^e$, and $A_v = \sqrt[4]{2}$ when $F_v \in \mathcal{M}^\phi$ (see footnote 6). Moreover, any $(\xi^{(0)}, \Sigma^{(0)})$ is an approximate zero of (19) as long as

$$\bar{\alpha}(\xi^{(0)}, \Sigma^{(0)}) := \frac{1}{IT}\left\|\left[\partial^2_{(\xi,\Sigma)}\mathscr{L}(\xi^{(0)}, \Sigma^{(0)})\right]^{-1} \partial_{(\xi,\Sigma)}\mathscr{L}(\xi^{(0)}, \Sigma^{(0)})\right\|$$
$$\times \left\|\left[\frac{\partial^2_{(\xi,\Sigma)}\mathscr{L}(\xi^{(0)}, \Sigma^{(0)})}{IT}\right]^{-1}\right\| \max\left\{1, \left\|\left[\frac{\partial^2_{(\xi,\Sigma)}\mathscr{L}(\xi^{(0)}, \Sigma^{(0)})}{IT}\right]^{-1}\right\|\right\}$$
$$< \alpha_0 \left(\max\left\{1, \frac{A_v J(e-1)\underline{\varsigma}^{-1}\bar{x}(a+1)^{1/(JT+KP)}}{(a+1)^{1/(JT+KP)}-1}\right\}\right)^{-3} \left(\frac{JT+KP}{3}\right)^{-3/2}. \quad \textbf{(20)}$$

When G is i.i.d. Gumbel or Gaussian (or a mixture of both), given y_{it}, because of Corollaries 1(b) and 1(c), $\prod_{j=1}^{J}\left[\sigma_{ij}(\xi_t + x_{it}\Sigma v_i; G)\right]^{\mathbf{1}\{y_{it}=j\}} = \sigma_{iy_{it}}(\xi_t + x_{it}\Sigma v_i; G)$ is real analytic with respect to (ξ_t, Σ) in their domains with a uniform radius of convergence and its higher-order derivatives are bounded by factorial rates

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

(as in Lemmas A2 and A3 in Appendix A) as long as $F_v \in \mathcal{F}^e$ (or $F_v \in \mathcal{M}^{f_{e+}}$ when $G \in \mathcal{M}^{\phi}$, with f_{e+} the density corresponding to the distribution $F_{e+} \in \mathcal{F}^{e+}$). Consequently, $\int \prod_{j=1}^{J} \left[\sigma_{ij}(\xi_t + x_{it}\Sigma v_i; G)\right]^{\mathbf{1}\{y_{it}=j\}} dF_v(v_i)$ is still real analytic with a uniform radius of convergence and its higher-order derivatives are bounded by factorial rates. As shown in Lemma F1 in Appendix F, these two properties still hold after the log transformation in $\ell_{it}(\xi, \Sigma)$ and lead to Corollary 5 by following the same strategy as in the proof of Corollary 4.

Different from the demand inverse of (13), which has a unique solution, $\mathscr{L}(\xi, \Sigma)$ will often have multiple local maxima and therefore $\partial_{(\xi,\Sigma)}\mathscr{L}(\xi, \Sigma) = 0$ multiple solutions. Corollary 5 guarantees the practical usefulness of the NR method also in mitigating the consequences of such multiplicity. Thanks to the supra-exponential convergence rate of (19) (see discussion around (15)), one will be able to quickly back out all the solutions of (19) by performing the NR method multiple times for different starting values. For each of the starting values, one finds a local maximum by performing a few more iterations after having established that $(\xi^{(n)}, \Sigma^{(n)})$ is an approximate zero, using (20):

$$\left\|(\xi^{(n+1)}, \Sigma^{(n+1)}) - (\xi^{(n)}, \Sigma^{(n)})\right\|$$
$$< IT \frac{\alpha_0 \left(\max\left\{1, \frac{A_v J(e-1)\underline{\varsigma}^{-1}\bar{x}(a+1)^{1/(JT+KP)}}{(a+1)^{1/(JT+KP)}-1}\right\}\right)^{-3} \left(\frac{JT+KP}{3}\right)^{-3/2}}{\left\|\left[\frac{\partial^2_{(\xi,\Sigma)}\mathscr{L}(\xi^{(n)}, \Sigma^{(n)})}{IT}\right]^{-1}\right\| \max\left\{1, \left\|\left[\frac{\partial^2_{(\xi,\Sigma)}\mathscr{L}(\xi^{(n)}, \Sigma^{(n)})}{IT}\right]^{-1}\right\|\right\}}. \quad \textbf{(21)}$$

Once all solutions of (19) are recovered, one can finally select the global maximum (i.e., the MLE) as the solution corresponding to the largest value of the log-likelihood function.

6. MONTE CARLO SIMULATIONS

In this section, we perform Monte Carlo simulations to illustrate the practical uses of Corollary 4 in implementing demand inverses in the context of the BLP approach.[18] As documented by Conlon and Gortmaker (2020) in extensive numerical experiments, Newton-type methods can be more effective than classic FP algorithms to implement demand inverses. While Conlon and Gortmaker (2020) focus on starting values in the basin of attraction of the unique solution (i.e., "close" starting values),[19] it is however possible for NR algorithms not to converge when the starting values are "distant" from the unique solution. In contrast, as discussed in Section 5.1, while slower in the proximity of the unique solution, the

[18] The MATLAB code of the simulations can be downloaded from the authors' websites.

[19] Conlon and Gortmaker (2020) use the vector of log-shares $\left(\log(p_j) - \log(p_0)\right)_{j=1}^{J}$, where $p_0 = 1 - \sum_j p_j$, as starting values. This would be the unique solution of the demand system in the case of a multinomial logit model (without random coefficients). Their numerical results illustrate that these conventional starting values are "close" to the unique solution, in the sense that the Newton-type methods they implement always converge to it.

convergence of FP algorithms based on contraction mappings is guaranteed from *any* starting value. This suggests that HAs along the lines of Algorithm 1 may be more effective in the case of "distant" starting values.

We investigate these possibilities relying on a data generating process along the lines of Conlon and Gortmaker (2020). In a first set of experiments, we study the role of approximate zeros in the numerical performance of FP algorithms based on contraction mappings versus NR algorithm (14) in the case of "close" starting values (i.e., close to the unique solution). In a second set of experiments, we then investigate the relative performance of FP algorithms based on contraction mappings versus HAs (Algorithm 1) in the case of "distant" starting values (i.e., starting values from which NR algorithms alone typically do not converge). In what follows we summarize the results of our numerical experiments, while we report the details of the data generating process, the generation of "close" and "distant" starting values, and the formulation of the FP algorithm in Appendix G.

On the one hand, Corollary 4 implies that NR algorithm (14) converges to the unique solution at supra-exponential rate after reaching an approximate zero. However, from a starting value that is "close" to the unique solution, the theory is silent regarding the number of iterations required to reach an approximate zero. On the other hand, FP algorithms based on contraction mappings are ensured to achieve exponential convergence from any starting value. As a consequence, if (14) takes a short time to reach an approximate zero, the overall convergence time could be shorter than that of a FP algorithm. Otherwise, if (14) takes a long time to reach an approximate zero, even in the case of "close" starting values, FP algorithms could be numerically more convenient.

Table 1 reports the execution times of FP and of NR algorithms in the case of "close" starting values. We study scenarios with different choice set sizes ($J = 25,\ 50$, and 100) and with or without a random coefficient on price. For each scenario and algorithm, Table 1 summarizes the time needed to reach an approximate zero (i.e., to satisfy condition (17)) from a "close" starting value, the time needed for convergence from an approximate zero, and the total of the two. All times are measured in seconds and are computed as the average among 200 randomly drawn starting values within a close distance from the unique solution. In line with the numerical findings by Conlon and Gortmaker (2020), from "close" starting values the NR algorithm is much faster than the FP algorithm to achieve numerical convergence in all scenarios considered: between 6 and 8.5 times faster and, in relative terms, faster for larger demand systems (larger J). Decomposing the total execution times, we find both that the NR algorithm is faster in reaching an approximate zero than the FP algorithm and that, after having reached an approximate zero, the NR algorithm greatly accelerates its speed of convergence while the FP algorithm does not.

When the starting values are "distant" from the unique solution, the NR algorithm may take a long time to converge or even fail to converge. In such cases, FP algorithms still deliver numerical convergence in theory, but potentially at a slow pace. As discussed in Section 5.1, some existing papers (e.g., Rust, 1987;

TABLE 1. Fixed Point (FP) versus Newton–Raphson (NR): "Close" Starting Values

No. of products	$J = 25$		$J = 50$		$J = 100$	
Algorithm	FP	NR	FP	NR	FP	NR
	No random coefficient on price					
Execution time (s)						
Total time	0.1542	0.0248	0.2567	0.0368	0.3741	0.0441
Time before approx. zero	0.0643	0.0203	0.1193	0.0309	0.1985	0.0379
Time after approx. zero	0.0899	0.0045	0.1374	0.0059	0.1756	0.0062
	Random coefficient on price					
Execution time (s)						
Total time	0.1506	0.0243	0.2632	0.0358	0.3696	0.0449
Time before approx. zero	0.0622	0.0200	0.1220	0.0303	0.1948	0.0385
Time after approx. zero	0.0884	0.0043	0.1412	0.0055	0.1748	0.0064

Note: Step tolerance level is 10^{-14}. Average statistics over 200 starting values randomly drawn within a distance of $[\delta_j - 1, \delta_j + 1]$ along each dimension j around the unique solution $(\delta_j)_{j=1}^{J}$. Max number of iterations is $1,000$. See Appendix G for specification of data generating process and further details.

Iskhakov et al., 2016) propose HAs (such as Algorithm 1) that combine a FP step with a NR step, so to benefit from the relative strengths of both procedures. While, in general, it is unclear after how many iterations of the FP step to switch to the NR step, Corollary 4 suggests to switch to the NR step after having reached an approximate zero, that is, when, during the FP step, condition (18) is satisfied. Note that, because not all points in the basin of attraction of the unique solution need to be approximate zeros, switching rule (18) is sufficient but may not be necessary for an efficient switch to the NR step. To explore this possibility, we also implement a more lenient switching rule than (18) that replaces Δ_{H}^{*} by the larger threshold $\sqrt[3]{\Delta_{\mathrm{H}}^{*}}$.

Table 2 reports the execution times of the FP algorithm and the HA (Algorithm 1 in Section 5.1) in the case of "distant" starting values. We study the same scenarios as in Table 1 and decompose execution times into its FP and NR components (for the case of the "pure" FP algorithm, the total execution time only corresponds to the "FP time"). All times are measured in seconds and are computed as the average among 200 randomly drawn starting values within a large distance from the unique solution.[20] We use two different switching rules to implement the HA, one based on the sufficient condition for an approximate zero, $\left\| \delta^{(n+1)} - \delta^{(n)} \right\| < \Delta_{\mathrm{H}}^{*}$, (denoted by HA Δ_{H}^{*}) and the other based on the more lenient $\left\| \delta^{(n+1)} - \delta^{(n)} \right\| < \sqrt[3]{\Delta_{\mathrm{H}}^{*}}$ (denoted by HA $\sqrt[3]{\Delta_{\mathrm{H}}^{*}}$).

[20] With these "distant" starting values, in more than 90% of cases the NR algorithm alone fails to converge, stressing the practical importance of the FP step to initialize the NR step.

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

Table 2. Fixed Point (FP) versus Hybrid Algorithm (HA): "Distant" Starting Values

No. of products	$J=25$			$J=50$			$J=100$		
Algorithm	FP	HA Δ_{H}^{*}	HA $\sqrt[3]{\Delta_{\mathrm{H}}^{*}}$	FP	HA Δ_{H}^{*}	HA $\sqrt[3]{\Delta_{\mathrm{H}}^{*}}$	FP	HA Δ_{H}^{*}	HA $\sqrt[3]{\Delta_{\mathrm{H}}^{*}}$
				No random coefficient on price					
Execution time (s)									
Total time	0.1597	0.0804	0.0434	0.2745	0.1436	0.0743	0.3910	0.2245	0.1175
FP time	0.1597	0.0702	0.0300	0.2745	0.1309	0.0568	0.3910	0.2123	0.0923
NR time	–	0.0102	0.0134	–	0.0127	0.0175	–	0.0122	0.0252
				Random coefficient on price					
Execution time (s)									
Total Time	0.1664	0.0805	0.0471	0.2848	0.1453	0.0721	0.4049	0.2204	0.1133
FP time	0.1664	0.0709	0.0328	0.2848	0.1328	0.0554	0.4049	0.2086	0.0893
NR time	–	0.0096	–	0.0143	0.0125	0.0167	–	0.0118	0.0240

Note: Step tolerance level is 10^{-14}. Average statistics over 200 starting values randomly drawn within a distance of $[\delta_j-5,\delta_j+5]$ along each dimension j around the unique solution $(\delta_j)_{j=1}^{J}$. The HA (Algorithm 1 in Section 5.1) is implemented using two switching rules: "HA Δ_{H}^{*}" switches from the FP step to the NR step when an approximate zero is reached ($\left\|\delta^{(n+1)}-\delta^{(n)}\right\|<\Delta_{\mathrm{H}}^{*}$), while "HA $\sqrt[3]{\Delta_{\mathrm{H}}^{*}}$" is more lenient and switches from the FP step to the NR step when $\left\|\delta^{(n+1)}-\delta^{(n)}\right\|<\sqrt[3]{\Delta_{\mathrm{H}}^{*}}$. Max number of iterations is 1,000. See Appendix G for specification of data generating process and further details.

Across all scenarios considered, as expected, both FP and HA always converge to the unique solution despite the "distant" starting values. Moreover, HA Δ_{H}^{*} tends to converge to the unique solution in approximately half the time needed by FP, and in turn, HA $\sqrt[3]{\Delta_{\mathrm{H}}^{*}}$ tends to require half the time needed by HA Δ_{H}^{*}. This suggests that switching to the NR step guarantees large time savings when an approximate zero is identified, but also that switching rule (18) may be "too" stringent. Increasing the switching threshold from Δ_{H}^{*} to $\sqrt[3]{\Delta_{\mathrm{H}}^{*}}$ leads to significant time savings, implying that efficient switches to the NR step can be implemented prior to satisfying (18), in that the more lenient switching threshold $\sqrt[3]{\Delta_{\mathrm{H}}^{*}}$ already identifies the basin of attraction of the unique solution.

6.1. Practical Suggestions for Numerical Implementation

Taken together, the theoretical results from the previous section and the Monte Carlo simulations highlight some practical recommendations useful to speed up the numerical implementation of demand inverses in the BLP approach:

- To start with, one can use the vector of log-shares $\left(\log(p_j)-\log(p_0)\right)_{j=1}^{J}$ as starting values (see footnote 19) and a NR algorithm along the lines of (14) with (i) step tolerance level of 10^{-14} and (ii) 1,000 as maximal number of iterations.
- When the NR algorithm (14) does not converge within the maximal number of iterations, this is evidence of considerable unobserved heterogeneity and that the multinomial logit starting values may be "distant" from the unique solution.

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

- In this case, one can still rely on the same starting values $\left(\log(p_j) - \log(p_0)\right)_{j=1}^{J}$ but use an HA along the lines of Algorithm 1 with (i) step tolerance level of 10^{-14}, (ii) 1,000 as maximal number of iterations, and (iii) switching rule between the FP step and the NR step along the lines of $\sqrt[3]{\Delta_{\mathrm{H}}^{*}}$ (with $\Delta_{\mathrm{H}}^{*} = \Delta^{*} p_{0t}$ and Δ^{*} as defined in (17)).
- If the procedure at the previous bullet point fails to converge, the lenient switching rule $\sqrt[3]{\Delta_{\mathrm{H}}^{*}}$ is then too loose. In this case, we suggest to tighten the switching rule between the FP step and the NR step to Δ_{H}^{*} so that numerical convergence is guaranteed (at the cost of a longer execution time).

7. CONCLUSIONS

We consider a class of discrete choice models of demand with an index structure in the indirect utilities, with any distribution of random coefficients, and which can include endogenous regressors. In the first part of the paper, we demonstrate that any model in this class can be approximated uniformly and arbitrarily well by a real analytic demand model. In the second part of the paper, we discuss the econometric advantages of real analytic demand models in terms of nonparametric and semi-nonparametric identification, extrapolation to hypothetical counterfactuals, numerical implementation of demand inverses in the context of aggregate market-level data, and numerical implementation of the MLE in the context of disaggregate individual-level data.

On the one hand, these results are encouraging as Theorem 1 and Corollary 1 illustrate that the class of real analytic demand models is relevant in practice, in that—to the best of our knowledge—most empirical papers dealing with demand estimation in applications with more than a few products typically specify mixed logit or mixed probit models. In addition, for any other demand model which is not real analytic, Theorem 2 shows that there exists a real analytic counterpart that can approximate it uniformly and arbitrarily well.

On the other hand, however, these results are subject to at least two caveats whose investigation we leave to future research. First, Theorem 2 is an "existence" or "possibility" result which does not provide practical guidance on how to select the real analytic approximants: they exist and they have the general mixed logit and mixed probit forms described in the theorem, but we may not know how to specify them in concrete examples. Second, the econometric advantages discussed in the paper apply directly to the real analytic demand models described in Theorem 1 and Corollary 1 and to the real analytic approximants in Theorem 2, but not necessarily to demand models that are not real analytic. Lack of real analyticity can occur, for example, when the distribution function G has jumps, so that the resulting demand function is discontinuous (Smith, 1935). Or again when the distribution function G has kinks, so that the resulting demand function is non-differentiable (even though potentially continuous). In such cases, Theorem 2 is silent on whether any of the properties of the real analytic approximant carries through the underlying demand model targeted by the approximation.

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

While this paper studies some of the econometric advantages of real analyticity for a class of static demand models along the lines of Berry et al. (1995), future research could extend the investigation to other popular classes of economic models. In particular, the estimation of many structural models relies on FP inner loops which are similar in nature to solving demand inverses in the BLP approach. For example, dynamic demand models (Rust, 1987; Gowrisankaran and Rysman, 2012) and (static and dynamic) entry models with incomplete information (Seim, 2006; Aguirregabiria and Mira, 2007) share this common feature. On the one hand, solving such FP inner loops typically gives rise to numerical problems similar to those faced in the context of demand inverses and sometimes even worse due to additional complexities such as large state spaces or multiplicity of equilibria. On the other hand, however, many of these structural models are real analytic and their practical implementation could be greatly simplified thanks to numerical advantages similar to those presented in this paper.

APPENDIX

To simplify exposition, we refer to $F(\cdot;d_i)$ simply as F and suppress from the notation d_i, t, and G, when these are all held constant. We also suppress G from the notation of the radius of convergence $r(\xi_t;F,G)$. Throughout the Appendix, we use the following inequality that follows from the multinomial theorem:

$$\binom{L+J-1}{L} = \sum_{\sum_{r=1}^{J} l_r = L} (1) \leq \sum_{\sum l_r = L} \frac{L!}{\prod_{r=1}^{J} l_r!} = J^L. \qquad \textbf{(A.1)}$$

A. Proof of Theorem 1

Proof of Statement (a), Mixed Logit. When G is i.i.d. Gumbel, we obtain

$$\sigma_j(\xi;X,F) = \int \frac{\exp\{\xi_j + g_j(x_j;\beta_i)\}}{1+\sum_{r=1}^{J}\exp\{\xi_r + g_r(x_r;\beta_i)\}} dF(\beta_i).$$

For simplicity, we denote $\sigma_{ij} = \frac{\exp\{\xi_j+g_j(x_j;\beta_i)\}}{1+\sum_{r=1}^{J}\exp\{\xi_r+g_r(x_r;\beta_i)\}}$. The key of the proof is to bound the higher-order derivatives of σ_{ij} with respect to ξ, $\frac{\partial^l \sigma_{ij}}{\prod_{r=1}^{J} \partial \xi_r^{l_r}}$, where l is an integer and $\sum_{r=1}^{J} l_r = l$. This is achieved by the following two lemmas.

LEMMA A1. *For any nonnegative integer l,*

$$\sup_{\xi,\beta_i,j,r} \left| \frac{\partial^l \sigma_{ij}}{\partial \xi_r^l} \right| \leq A_l l!,$$

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

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

Proof. Define $a_l = \sup_{\xi,\beta_i,j,r} \left| \frac{\partial^l \sigma_{ij}}{\partial \xi_j^l} \right|$ and note that

$$\exp\{\xi_j + g_j(x_j;\beta_i)\} = \sigma_{ij} \times \left(1 + \sum_{r=1}^{J} \exp\{\xi_r + g_r(x_r;\beta_i)\}\right),$$

$$\mathbf{1}\{r=j\}\exp\{\xi_j + g_j(x_j;\beta_i)\} = \frac{\partial^l \exp\{\xi_j + g_j(x_j;\beta_i)\}}{\partial \xi_r^l}$$

$$= \frac{\partial^l \left[\sigma_{ij} \times \left(1 + \sum_{s=1}^{J} \exp\{\xi_s + g_s(x_s;\beta_i)\}\right)\right]}{\partial \xi_r^l}$$

$$= \sum_{k=0}^{l} \binom{l}{k} \frac{\partial^k \sigma_{ij}}{\partial \xi_r^k} \times \frac{\partial^{l-k}\left(1 + \sum_{s=1}^{J} \exp\{\xi_s + g_s(x_s;\beta_i)\}\right)}{\partial \xi_r^{l-k}}$$

$$= \frac{\partial^l \sigma_{ij}}{\partial \xi_r^l}\left(1 + \sum_{r=1}^{J} \exp\{\xi_r + g_r(x_r;\beta_i)\}\right) + \sum_{k=0}^{l-1} \binom{l}{k} \frac{\partial^k \sigma_{ij}}{\partial \xi_r^k} \exp\{\xi_r + g_r(x_r;\beta_i)\},$$

$$\frac{\partial^l \sigma_{ij}}{\partial \xi_r^l} = \sigma_{ir} \times \left(\mathbf{1}\{r=j\} - \sum_{k=0}^{l-1} \binom{l}{k} \frac{\partial^k \sigma_{ij}}{\partial \xi_r^k}\right),$$

$$\left|\frac{\partial^l \sigma_{ij}}{\partial \xi_r^l}\right| \leq 1 + \sum_{k=0}^{l-1} \binom{l}{k} \left|\frac{\partial^k \sigma_{ij}}{\partial \xi_r^k}\right| \leq 1 + \sum_{k=0}^{l-1} \binom{l}{k} a_k \tag{A.2}$$

for any $j, r = 1, \ldots, J$. Then,

$$a_l \leq 1 + \sum_{k=0}^{l-1} \binom{l}{k} a_k,$$

$$\frac{a_l}{l!} \leq \frac{1}{l!} + \sum_{k=0}^{l-1} \frac{a_k}{k!} \frac{1}{(l-k)!}.$$

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 $a_1 = \sup_{\xi,\beta_i,j,r} \left|\frac{\partial \sigma_{ij}}{\partial \xi_r}\right| = \sup_{\xi,\beta_i,j,r} \{\sigma_{ij}(1-\sigma_{ij}), \sigma_{ij}\sigma_{ir}\} \leq 1 < e = A_1$. Suppose that $\frac{a_k}{k!} \leq A_k$ holds for $k = 1, \ldots, l-1$. Note that $A_l = \frac{1}{l!} + (e-1)A_{l-1} > A_{l-1}$, for any $l \geq 0$. Then,

$$\frac{a_l}{l!} \leq \frac{1}{l!} + \sum_{k=0}^{l-1} \frac{a_k}{k!} \frac{1}{(l-k)!}$$

$$\leq \frac{1}{l!} + \sum_{k=0}^{l-1} A_k \frac{1}{(l-k)!}$$

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

$$
\begin{aligned}
&\leq \frac{1}{l!} + A_{l-1} \sum_{k=0}^{l-1} \frac{1}{(l-k)!} \\
&\leq \frac{1}{l!} + A_{l-1}(e-1) \\
&= A_l. \qquad \textbf{(A.3)}
\end{aligned}
$$

As a consequence, the inequality holds for any $l > 0$ and $a_l = \sup_{\xi,\beta_i,j,r} \left| \frac{\partial^l \sigma_{ij}}{\partial \xi_r^l} \right| \leq A_l l!$. This completes the proof of Lemma A1. □

The next lemma controls $\frac{\partial^l \sigma_{ij}}{\prod_{r=1}^J \partial \xi_r^{l_r}}$.

LEMMA A2. *For any* $j = 1, \ldots, J$, $l \geq 0$, *and* $\sum_{r=1}^J l_r = l$,

$$
\sup_{\xi \in \mathbb{R}^J, \beta_i \in \mathbb{R}^K} \left| \frac{\partial^l \sigma_{ij}}{\prod_{r=1}^J \partial \xi_r^{l_r}} \right| \leq e^{1/(e-1)} (e-1)^l \prod_{r=1}^J l_r!.
$$

Proof. We prove the result by induction. For $l = 0$, the result holds trivially. For $l = 1$, the result follows directly from Lemma A1 with $l = 1$. Suppose that $l = 2$. When $l_r = 2$, according to Lemma A1, we have $\left| \frac{\partial^2 \sigma_{ij}}{\partial \xi_j^2} \right| \leq A_2 2!$. Note that $A_2 = (e-1)^2 \left(1 + \frac{1}{e-1} + \frac{1}{2(e-1)^2}\right) \leq e^{1/(e-1)}(e-1)^2$. For $l_r = l_s = 1$, $r \neq s$,

$$
\exp\{\xi_j + g_j(x_j;\beta_i)\} = \sigma_{ij} \times \left(1 + \sum_{r=1}^J \exp\{\xi_r + g_r(x_r;\beta_i)\}\right),
$$

$$
0 = \frac{\partial^2 \sigma_{ij}}{\partial \xi_r \partial \xi_s} \left(1 + \sum_{r=1}^J \exp\{\xi_r + g_r(x_r;\beta_i)\}\right) + \frac{\partial \sigma_{ij}}{\partial \xi_r} \exp\{\xi_s + g_s(x_s;\beta_i)\} + \frac{\partial \sigma_{ij}}{\partial \xi_s} \exp\{\xi_r + g_r(x_r;\beta_i)\},
$$

$$
\frac{\partial^2 \sigma_{ij}}{\partial \xi_r \partial \xi_s} = -\sigma_{is} \frac{\partial \sigma_{ij}}{\partial \xi_r} - \sigma_{ir} \frac{\partial \sigma_{ij}}{\partial \xi_s}.
$$

By using $\left| \frac{\partial \sigma_{ij}}{\partial \xi_r} \right|, \left| \frac{\partial \sigma_{ij}}{\partial \xi_s} \right| \leq 1$ and $\sigma_{is} + \sigma_{ir} < 1$, we have $\sup_{\xi,\beta_i} \left| \frac{\partial^2 \sigma_{ij}}{\partial \xi_r \partial \xi_s} \right| \leq 1 < e^{1/(e-1)}(e-1)^2$. As a consequence, the conclusion holds for $l = 2$.

Suppose that for $k = 0, \ldots, l-1$ the inequality holds for any $\sum_{r=1}^J l_r = k$. First, remember that $A_l = (e-1)^l \sum_{k=0}^l \frac{1}{(e-1)^k k!}$, as defined in Lemma A1, is smaller than $e^{1/(e-1)}(e-1)^l$. Hence, the conclusion holds for any $l > 0$ with $l_r = l$. It remains to show that the conclusion holds also when $l_r, l_s > 0$, for some $r \neq s$.

By taking l_rth derivatives of both sides of the first equation in (A.2) with respect to ξ_r, we obtain

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

$$\mathbf{1}\{r=j\}\exp\{\xi_j+g_j(x_j;\beta_i)\} = \frac{\partial^{l_r}\sigma_{ij}}{\partial \xi_r^{l_r}}\left(1+\sum_{s=1}^{J}\exp\{\xi_s+g_s(x_s;\beta_i)\}\right) + \sum_{k=0}^{l_r-1}\binom{l_r}{k}\frac{\partial^k\sigma_{ij}}{\partial\xi_r^k}\exp\{\xi_r+g_r(x_r;\beta_i)\}. \quad \textbf{(A.4)}$$

Note that, by taking derivatives of both sides of equation (A.4) with respect to ξ_s, $s\neq r$, the left-hand side vanishes. We then have

$$0=\frac{\partial^{l_r+l_s}\sigma_{ij}}{\partial\xi_r^{l_r}\partial\xi_s^{l_s}}\left(1+\sum_{m=1}^{J}\exp\{\xi_m+g_m(x_m;\beta_i)\}\right) + \exp\{\xi_s+g_s(x_s;\beta_i)\}\sum_{k=0}^{l_s-1}\binom{l_s}{k}\frac{\partial^{l_r+k}\sigma_{ij}}{\partial\xi_r^{l_r}\partial\xi_s^{k}} + \exp\{\xi_r+g_r(x_r;\beta_i)\}\sum_{k=0}^{l_r-1}\binom{l_r}{k}\frac{\partial^{k+l_s}\sigma_{ij}}{\partial\xi_r^{k}\partial\xi_s^{l_s}}.$$

Finally, using $\sum_{r=1}^{J}\sigma_{ir}<1$, we obtain

$$0=\frac{\partial^{l}\sigma_{ij}}{\prod_{r=1}^{J}\partial\xi_r^{l_r}}\left(1+\sum_{m=1}^{J}\exp\{\xi_m+g_m(x_m;\beta_i)\}\right)+\sum_{s=1}^{J}\exp\{\xi_s+g_s(x_s;\beta_i)\}\sum_{k=0}^{l_s-1}\binom{l_s}{k}\frac{\partial^{l-l_s+k}\sigma_{ij}}{\partial\xi_s^{k}\prod_{m\neq s}\partial\xi_m^{l_m}},$$

$$\frac{\partial^{l}\sigma_{ij}}{\prod_{r=1}^{J}\partial\xi_r^{l_r}}=-\sum_{r=1}^{J}\sigma_{ir}\sum_{k=0}^{l_r-1}\binom{l_r}{k}\frac{\partial^{l-l_r+k}\sigma_{ij}}{\partial\xi_r^{k}\prod_{s\neq r}\partial\xi_s^{l_s}},$$

$$\left(\frac{1}{\prod_{r=1}^{J}l_r!}\right)\left(\frac{\partial^{l}\sigma_{ij}}{\prod_{r=1}^{J}\partial\xi_r^{l_r}}\right)=-\sum_{r=1}^{J}\sigma_{ir}\sum_{k=0}^{l_r-1}\left[\frac{1}{(l_r-k)!\,k!\prod_{s\neq r}l_s!}\right]\frac{\partial^{l-l_r+k}\sigma_{ij}}{\partial\xi_r^{k}\prod_{s\neq r}\partial\xi_s^{l_s}},$$

$$\left|\left(\frac{1}{\prod_{r=1}^{J}l_r!}\right)\left(\frac{\partial^{l}\sigma_{ij}}{\prod_{r=1}^{J}\partial\xi_r^{l_r}}\right)\right|\leq\max_{r=1,\ldots,J}\left\{\sum_{k=0}^{l_r-1}\frac{1}{(l_r-k)!}\left|\left(\frac{1}{k!\prod_{s\neq r}l_s!}\right)\left(\frac{\partial^{l-l_r+k}\sigma_{ij}}{\partial\xi_r^{k}\prod_{s\neq r}\partial\xi_s^{l_s}}\right)\right|\right\}. \quad \textbf{(A.5)}$$

Consequently, applying the conclusion for any $k=0,\ldots,l-1$ to the last inequality in (A.5), we obtain

$$\begin{aligned}\sup_{\xi,\beta_i}\left|\frac{1}{\prod_{r=1}^{J}l_r!}\frac{\partial^{l}\sigma_{ij}}{\prod_{r=1}^{J}\partial\xi_r^{l_r}}\right| &\leq e^{1/(e-1)}\max_{r=1,\ldots,J}\left\{\sum_{k=0}^{l_r-1}\frac{1}{(l_r-k)!}(e-1)^{l-l_r+k}\right\}\\ &=e^{1/(e-1)}(e-1)^{l}\max_{r=1,\ldots,J}\left\{\sum_{k=1}^{l_r}\frac{1}{k!}(e-1)^{-k}\right\}\\ &\leq e^{1/(e-1)}(e-1)^{l}(e^{(e-1)^{-1}}-1)\\ &< e^{1/(e-1)}(e-1)^{l}.\end{aligned} \quad \textbf{(A.6)}$$

Hence, $\sup_{\xi,\beta_i}\left|\frac{\partial^{l}\sigma_{ij}}{\prod_{r=1}^{J}\partial\xi_r^{l_r}}\right|\leq e^{1/(e-1)}(e-1)^{l}\prod_{r=1}^{J}l_r!$ and the conclusion holds for $\sum_{r=1}^{J}l_r=l$. This completes the proof of Lemma A2. □

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

Using Lemmas A1 and A2, we can express the Taylor series of $\sigma_j(\xi';X,F)$ at ξ as

$$\begin{aligned}\left|\sum_{L=0}^{\infty}\frac{1}{L!}\left[\sum_{r=1}^{J}(\xi'_r-\xi_r)\frac{\partial}{\partial\xi_r}\right]^L\sigma_j(\xi;X,F)\right| &\leq \sum_{L=0}^{\infty}\frac{1}{L!}d^L\sum_{\sum l_r=L}\frac{L!}{\prod_{r=1}^J l_r!}\left|\frac{\partial^L\sigma_j(\xi;X,F)}{\prod_{r=1}^J\partial\xi_r^{l_r}}\right| \\ &\leq e^{1/(e-1)}\sum_{L=0}^{\infty}\frac{1}{L!}d^L\sum_{\sum l_r=L}\frac{L!}{\prod_{r=1}^J l_r!}(e-1)^L\prod_{r=1}^J l_r! \\ &\leq e^{1/(e-1)}\sum_{L=0}^{\infty}(e-1)^Ld^L\sum_{\sum l_r=L}(1) \\ &\leq e^{1/(e-1)}\sum_{L=0}^{\infty}d^L[J(e-1)]^L, \qquad \textbf{(A.7)}\end{aligned}$$

where the last inequality results from (A.1) and $d=\|\xi'-\xi\|$. Consequently, whenever $d<d^*=\frac{1}{J(e-1)}$, the Taylor series (A.7) converges. Finally, given any $\tilde{d}<d^*$, by applying Taylor's theorem to the multivariate function $\sigma_j(\xi';X,F)$, we obtain that uniformly for $\|\xi'-\xi\|<\tilde{d}$,

$$\begin{aligned}&\left|\sigma_j(\xi';X,F)-\sum_{L=0}^{R}\frac{1}{L!}\left[\sum_{r=1}^{J}(\xi'_r-\xi_r)\frac{\partial}{\partial\xi_r}\right]^L\sigma_j(\xi;X,F)\right| \\ \leq&\tilde{d}^{R+1}\sum_{\sum l_r=R+1}\frac{1}{\prod l_r!}\sup_{\|\xi'-\xi\|<\tilde{d}}\left|\frac{\partial^{R+1}\sigma_j(\xi';X,F)}{\prod_{r=1}^J\partial\xi_r^{l_r}}\right| \\ \leq&e^{1/(e-1)}(\tilde{d}(e-1)J)^{R+1} \\ \to&0\end{aligned}$$

as $R\to\infty$. To conclude, the Taylor series of $\sigma_j(\xi';X,F)$ at ξ converges to $\sigma_j(\xi';X,F)$ and $\sigma_j(\xi';X,F)$ is therefore real analytic with respect to ξ' at ξ_t. Finally, the radius of convergence $r(\xi_t;F)$ is at least $d^*=\frac{1}{J(e-1)}$ and only depends on J but not on ξ_t and F. As a result, $\inf_{\xi_t\in\mathbb{R}^J,F}r(\xi_t;F)>0$. This completes the proof of Theorem 1, statement (a).

Proof of Statement (b), Mixed Probit. Without loss of generality, suppose that $j=1$ and $\tilde{\varepsilon}=(\varepsilon_{i0}-\varepsilon_{i1},\varepsilon_{i2}-\varepsilon_{i1},\dots,\varepsilon_{iJ}-\varepsilon_{i1})$ follows a centered multivariate Gaussian distribution with a positive-definite variance–covariance matrix.[21] Denote by $\Sigma\Sigma^{\mathrm{T}}$ the unique Cholesky decomposition of the variance–covariance matrix with $\Sigma>0$. It then follows that

$$\sigma_1(\xi;X,F)=\kappa(\tilde{\xi}):=\int\prod_{j=1}^{J}\Phi\left(\tilde{\xi}_j+\tilde{g}(x_j;\beta_i)\right)dF(\beta_i),$$

[21] One can include the mean of ε_{ij} as a constant in the definition of g_j such that each ε_{ij} is centered. Moreover, the variance–covariance matrix of $\tilde{\varepsilon}$ is positive-definite as long as $(\varepsilon_{i0},\dots,\varepsilon_{iJ})$ is a nondegenerate Gaussian random vector.

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

where Φ denotes the standard normal distribution function and

$$\tilde{\xi} = \Sigma^{-1}(\xi_1, \xi_1 - \xi_2, \ldots, \xi_1 - \xi_J)^{\mathrm{T}},$$
$$\tilde{g}(x;\beta_i) = \Sigma^{-1}(g_1(x_1;\beta_i), g_1(x_1;\beta_i) - g_2(x_2;\beta_i), .., g_1(x_1;\beta_i) - g_J(x_J;\beta_i))^{\mathrm{T}}.$$

Proving statement (b) of Theorem 1 is equivalent to proving that $\kappa(\tilde{\xi})$ is real analytic with respect to $\tilde{\xi} \in \mathbb{R}^J$ and the radius of convergence is $+\infty$ at any $\tilde{\xi}$. To start, we first prove the following lemma.

LEMMA A3. *For $\sum_{j=1}^{J} l_j = L$, we have*

$$\left| \frac{\partial^L \kappa(\tilde{\xi})}{\prod_{j=1}^{J} \partial \tilde{\xi}_j^{l_j}} \right| \le (C/\sqrt{2\pi})^J \prod_{j=1}^{J} \sqrt{l_j!},$$

where C is a constant defined in (A.11).

Proof. First, we prove the statement for $l_j \ge 1$ for any $j = 1, \ldots, J$. Denote by ϕ the standard normal density function. Note that

$$\begin{aligned}
\frac{\partial^L \kappa(\tilde{\xi})}{\prod_{j=1}^{J} \partial \tilde{\xi}_j^{l_j}} &= \int \prod_{j=1}^{J} \frac{\partial^{l_j - 1} \phi(\tilde{\xi}_j + \tilde{g}_j(x_j;\beta_i))}{\partial \tilde{\xi}_j^{l_j - 1}} dF(\beta_i) \\
&= \pi^{-\frac{J}{2}} (-1)^{J+L} 2^{-\frac{L}{2}} \int \prod_{j=1}^{J} \mathrm{H}_{l_j - 1}\left(\frac{\tilde{\xi}_j + \tilde{g}_j(x_j;\beta_i)}{\sqrt{2}} \right) \exp\left\{ -\left(\frac{\tilde{\xi}_j + \tilde{g}_j(x_j;\beta_i)}{\sqrt{2}} \right)^2 \right\} dF(\beta_i) \\
&= \pi^{-\frac{J}{2}} (-1)^{J+L} 2^{-\frac{L}{2}} \int \prod_{j=1}^{J} \sqrt{ \mathrm{H}^2_{l_j - 1}\left(\frac{\tilde{\xi}_j + \tilde{g}_j(x_j;\beta_i)}{\sqrt{2}} \right) \exp\left\{ -\left(\frac{\tilde{\xi}_j + \tilde{g}_j(x_j;\beta_i)}{\sqrt{2}} \right)^2 \right\} } \\
&\quad \times \exp\left\{ -\frac{(\tilde{\xi}_j + \tilde{g}_j(x_j;\beta_i))^2}{4} \right\} dF(\beta_i),
\end{aligned}$$

where $\mathrm{H}_m(x) = (-1)^m \exp\{x^2\} \frac{d^m}{dx^m} \exp\{-x^2\}$ is the Hermite polynomial of order m. According to Theorem 1 in Krasikov (2004), for any $m \ge 6$, we have

$$\bar{H}_m^2 := \sup_{x \in \mathbb{R}} \mathrm{H}_m^2(x) \exp\{-x^2\} \le \frac{2C_m}{3(2m)^{1/6}} \exp\left\{ \frac{15}{8} \left[1 + \frac{12}{4(2m)^{1/3} - 9} \right] \right\}, \quad \textbf{(A.8)}$$

where

$$C_m = \begin{cases} \frac{2m\sqrt{4m-2}(m!)^2}{\sqrt{8m^2 - 8m + 3}((m/2)!)^2}, & \text{if } m \text{ is even,} \\ \frac{\sqrt{16m^2 - 16m + 6}m!(m-1)!}{\sqrt{2m-1}[((m-1)/2)!]^2}, & \text{if } m \text{ is odd.} \end{cases}$$

Using Stirling's formula to approximate $m!$,

$$\sqrt{2\pi m}(m/e)^m \exp\left\{ \frac{1}{12m+1} \right\} < m! < \sqrt{2\pi m}(m/e)^m \exp\left\{ \frac{1}{12m} \right\}, \quad \textbf{(A.9)}$$

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

we can bound C_m by $C_m \le E_m \times 2^m m!$, where

$$E_m = \sqrt{\frac{2}{\pi}} \exp\left\{\frac{1}{12(m-1)(6m-5)}\right\} \max\left\{\sqrt{\frac{8m^2-8m+3}{8m^2-12m+4}}, \sqrt{\frac{8m^2-4m}{8m^2-8m+3}}\right\}.$$

Then, for $m \ge 6$, we have

$$\bar{H}_m \le \sqrt{\frac{2E_m}{3(2m)^{1/6}}} \exp\left\{\frac{15}{16}\left(1+\frac{12}{4(2m)^{1/3}-9}\right)\right\} \times 2^{m/2}\sqrt{m!}.$$

While for $m = 0, 1, \dots, 5$, we have

$$\bar{H}_m \le \max_{0\le m\le 5}\left\{\frac{\bar{H}_m}{2^{m/2}\sqrt{m!}}\right\} \times 2^{m/2}\sqrt{m!}.$$

Note that $\max_{0\le m\le 5}\left\{\frac{\bar{H}_m}{2^{m/2}\sqrt{m!}}\right\}$ is finite. Moreover, $\sqrt{\frac{2E_m}{3(2m)^{1/6}}} \exp\left\{\frac{15}{16}\left(1+\frac{12}{4(2m)^{1/3}-9}\right)\right\}$ is decreasing in m and therefore bounded by its value at $m = 6$. We can then write

$$\bar{H}_m = \sup_{x\in\mathbb{R}} |\mathrm{H}_m(x)| \exp\{-x^2/2\} \le C \times 2^{m/2}\sqrt{m!}, \tag{A.10}$$

where

$$C = \max\left\{\sqrt{\frac{2E_6}{3(12)^{1/6}}} \exp\left\{\frac{15}{16}\left(1+\frac{12}{4\times(12)^{1/3}-9}\right)\right\}, \max_{0\le m\le 5}\left\{\frac{\bar{H}_m}{2^{m/2}\sqrt{m!}}\right\}\right\}. \tag{A.11}$$

As a result, plugging (A.10) in $\frac{\partial^L \kappa(\tilde{\xi})}{\prod_{j=1}^J \partial \tilde{\xi}_j^{l_j}}$, we obtain

$$\begin{aligned}\left|\frac{\partial^L \kappa(\tilde{\xi})}{\prod_{j=1}^J \partial \tilde{\xi}_j^{l_j}}\right| &\le (C/\sqrt{2\pi})^J \prod_{j=1}^J \sqrt{(l_j-1)!} \int \prod_{j=1}^J \exp\left\{-\frac{(\tilde{\xi}_j + \tilde{g}_j(x_j;\beta_i))^2}{4}\right\} dF(\beta_i)\\ &\le (C/\sqrt{2\pi})^J \prod_{j=1}^J \sqrt{l_j!}.\end{aligned}$$

Second, without loss of generality, suppose there exists $j = 1, \dots, k$ such that $l_j = 0$. Then,

$$\begin{aligned}\frac{\partial^L \kappa(\tilde{\xi})}{\prod_{j=1}^J \partial \tilde{\xi}_j^{l_j}} &= \int \prod_{j=1}^k \Phi(\tilde{\xi}_j + \tilde{g}_j(x_j;\beta_i)) \prod_{j=k+1}^J \frac{\partial^{l_j-1}\phi(\tilde{\xi}_j + \tilde{g}_j(x_j;\beta_i))}{\partial \tilde{\xi}_j^{l_j-1}} dF(\beta_i)\\ &= \pi^{-\frac{J-k}{2}}(-1)^{J-k+L} 2^{-\frac{L-k}{2}} \int \prod_{j=1}^k \Phi(\tilde{\xi}_j + \tilde{g}_j(x_j;\beta_i))\\ \times \prod_{j=k+1}^J &\sqrt{\mathrm{H}_{l_j-1}^2\left(\frac{\tilde{\xi}_j + \tilde{g}_j(x_j;\beta_i)}{\sqrt{2}}\right) \exp\left\{-\left(\frac{\tilde{\xi}_j + \tilde{g}_j(x_j;\beta_i)}{\sqrt{2}}\right)^2\right\}} \exp\left\{-\frac{(\tilde{\xi}_j + \tilde{g}_j(x_j;\beta_i))^2}{4}\right\} dF(\beta_i).\end{aligned}$$

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

Using the same arguments as above, $C \geq \sqrt{2\pi}$, $\Phi \leq 1$, and $k > 0$. We then obtain

$$\left| \frac{\partial^L \kappa(\tilde{\xi})}{\prod_{j=1}^J \partial \tilde{\xi}_j^{l_j}} \right| \leq (C/\sqrt{2\pi})^J \prod_{j=1}^J \sqrt{l_j!}.$$

The proof is completed. □

As in the derivation of inequality (A.7), we can bound the Taylor series of $\kappa(\tilde{\xi}')$ at $\tilde{\xi}$:

$$\begin{aligned}
\left| \sum_{L=0}^{\infty} \frac{1}{L!} \left[\sum_{r=1}^{J} (\tilde{\xi}_r' - \tilde{\xi}_r) \frac{\partial}{\partial \tilde{\xi}_r} \right]^L \kappa(\tilde{\xi}) \right| &\leq \sum_{L=0}^{\infty} \frac{1}{L!} d^L \sum_{\sum l_r = L} \frac{L!}{\prod_{r=1}^J l_r!} \left| \frac{\partial^L \kappa(\tilde{\xi})}{\prod_{r=1}^J \partial \tilde{\xi}_r^{l_r}} \right| \\
&\leq (C/\sqrt{2\pi})^J \sum_{L=0}^{\infty} d^L \sum_{\sum l_r = L} \frac{1}{\sqrt{\prod_{r=1}^J l_r!}} \\
&\leq (C/\sqrt{2\pi})^J \sum_{L=0}^{\infty} d^L \sqrt{\sum_{\sum l_r = L} \frac{1}{\prod_{r=1}^J l_r!} \sum_{\sum l_r = L} (1)} \\
&\leq (C/\sqrt{2\pi})^J \sum_{L=0}^{\infty} \frac{(\sqrt{J} d)^L}{\sqrt{L!}}, \qquad \textbf{(A.12)}
\end{aligned}$$

where $d = \|\tilde{\xi}' - \tilde{\xi}\|$. Note that the radius of convergence of the power series (as a function of d) on the right-hand side of (A.12) is $\lim_{L\to\infty} \left| \frac{(\sqrt{J})^L/\sqrt{L!}}{(\sqrt{J})^{L+1}/\sqrt{(L+1)!}} \right| = +\infty$. Consequently, the convergence of the Taylor series in (A.12) is always achieved. Then, for any $d^* > 0$ and uniformly for $\|\tilde{\xi}' - \tilde{\xi}\| < d^*$,

$$\left| \kappa(\tilde{\xi}') - \sum_{L=0}^{R} \frac{1}{L!} \left[\sum_{r=1}^{J} (\tilde{\xi}_r' - \tilde{\xi}_r) \frac{\partial}{\partial \tilde{\xi}_r} \right]^L \kappa(\tilde{\xi}) \right| \leq (C/\sqrt{2\pi})^J \frac{(\sqrt{J} d^*)^{R+1}}{\sqrt{(R+1)!}} \to 0$$

as $R \to \infty$. Therefore, the Taylor series of $\kappa(\tilde{\xi}')$ at $\tilde{\xi}$ converges to $\kappa(\tilde{\xi}')$ and $\kappa(\tilde{\xi}')$ is real analytic with respect to $\tilde{\xi}'$ at $\tilde{\xi}$ and the radius of convergence is $+\infty$. This completes the proof of Theorem 1, statement (b).

B. Proof of Corollary 1

In statements (a) and (b), we suppose that

$$\begin{aligned}
G(\varepsilon) &= r \sum_{i=1}^{m_1} c_{i1} \Phi_i(\varepsilon) + (1-r) \sum_{i=1}^{m_2} c_{i2} \Psi_i(\varepsilon) \\
&= r \sum_{i=1}^{m_1} c_{i1} \frac{1}{\varsigma_{i1}^{J+1}} \Phi\left(\frac{(\varepsilon_j)_{j=0}^J - \mu_{i1}}{\varsigma_{i1}} \right) + (1-r) \sum_{i=1}^{m_2} c_{i2} \frac{1}{\varsigma_{i2}^{J+1}} \Psi\left(\frac{(\varepsilon_j)_{j=0}^J - \mu_{i2}}{\varsigma_{i2}} \right),
\end{aligned} \qquad \textbf{(B.1)}$$

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

where $\Phi_i := \Phi\left(\frac{(\varepsilon_j)_{j=0}^J - \mu_{i1}}{\varsigma_{i1}}\right)$, $\Psi_i := \Psi\left(\frac{(\varepsilon_j)_{j=0}^J - \mu_{i2}}{\varsigma_{i2}}\right)$, and Φ and Ψ are, respectively, the distribution functions of i.i.d. standard Gaussian random variables and of i.i.d. Gumbel random variables in $\mathbb{R}^{J+1}$. Then,

$$\begin{aligned}\sigma_j(\xi;X,F,G) &= r\sum_{i=1}^{m_1} c_{i1}\sigma_j(\xi;X,F,\Phi_i) + (1-r)\sum_{i=1}^{m_2} c_{i2}\sigma_j(\xi;X,F,\Psi_i)\\ &= r\sum_{i=1}^{m_1} c_{i1}\sigma_j(\varsigma_{i1}^{-1}(\xi-\bar{\mu}_{i1});\varsigma_{i1}^{-1}X,F,\Phi) + (1-r)\sum_{i=1}^{m_2} c_{i2}\sigma_j(\varsigma_{i2}^{-1}(\xi-\bar{\mu}_{i2});\varsigma_{i2}^{-1}X,F,\Psi),\end{aligned}$$

where $\bar{\mu}_{i1} = (\mu_{ij1})_{j=1}^J - \mu_{i01}$ and $\bar{\mu}_{i2} = (\mu_{ij2})_{j=1}^J - \mu_{i02}$. Without loss of generality, we prove the corollary for $j=1$.

Proof of Statement (a). Because of Theorem 1, $\sigma_1(\xi;X,F,\Phi_i)$ and $\sigma_1(\xi;X,F,\Psi_i)$ are real analytic with respect to $\xi\in\mathbb{R}^J$ with radii of convergence uniformly bounded away from zero for $\xi\in\mathbb{R}^J$ and F. Then, $\sigma_j(\xi;X,F,G)$ is real analytic with respect to $\xi\in\mathbb{R}^J$ and $\inf\limits_{\xi\in\mathbb{R}^J,F} r(\xi;F) > 0$.

Proof of Statement (b). It suffices to prove that $\sigma_j(\xi;X,F,\Psi)$ and $\sigma_j(\xi;X,F,\Phi)$ are real analytic with respect to $(\xi,X)\in\mathbb{R}^J\times\mathbb{R}^{J\times K}$ with radii of convergence uniformly bounded away from zero for $(\xi,X)\in\mathbb{R}^J\times\mathbb{R}^{J\times K}$. For both cases, the proof is similar to that of Theorem 1 and the key is to bound $\frac{\partial^L\sigma_j(\xi;X,F,G)}{\prod_{j=1}^J \partial\xi_j^{l_{j0}}\prod_{1\le j\le J,1\le k\le K}\partial x_{jk}^{l_{jk}}}$ with $L=\sum_{j,k} l_{jk}$.

When $G=\Psi$, because of $g_j(x_{jt};\beta_i) = x_{jt}\beta_i$, we have

$$\frac{\partial^L\sigma_j(\xi;X,F,\Phi)}{\prod_{j=1}^J \partial\xi_j^{l_{j0}}\prod_{1\le j\le J,1\le k\le K}\partial x_{jk}^{l_{jk}}} = \int\prod_{k=1}^K \beta_{ik}^{\sum_{j=1}^J l_{jk}}\frac{\partial^L\sigma_{ij}}{\prod_{j=1}^J \partial\xi_j^{\sum_{k=0}^K l_{jk}}}dF(\beta_i).$$

Then, using Lemma A2 and $F\in\mathcal{F}^e$, we obtain

$$\left|\frac{\partial^L\sigma_j(\xi;X,F,\Phi)}{\prod_{j=1}^J \partial\xi_j^{l_{j0}}\prod_{1\le j\le J,1\le k\le K}\partial x_{jk}^{l_{jk}}}\right| \le e^{1/(e-1)}A^L(e-1)^L\prod_{j=1}^J\Big(\sum_{k=0}^K l_{jk}\Big)!,$$

where $L=\sum_{1\le j\le J,0\le k\le K} l_{jk}$ and the constant A is detailed in the definition of $\mathcal{F}^e$. Then, the Lth term in the Taylor series of $\sigma_j(\xi';X,F,\Psi)$ at ξ can be bounded as

$$\begin{aligned}&\left|\frac{1}{L!}\left[\sum_{r=1}^J(\xi_r'-\xi_r)\frac{\partial}{\partial\xi_r} + (x_{rk}'-x_{rk})\sum_{1\le r\le J,1\le k\le K}\frac{\partial}{\partial x_{rk}}\right]^L\sigma_j(\xi;X,F,\Phi)\right|\\ &\le\frac{d^L}{L!}\sum_{\sum_{1\le r\le J,0\le k\le K} l_{rk}=L}\frac{L!}{\prod_{1\le r\le J,0\le k\le K} l_{jk}!}\left|\frac{\partial^L\sigma_j(\xi;X,F,\Phi)}{\prod_{r=1}^J\partial\xi_r^{l_{r0}}\prod_{1\le r\le J,1\le k\le K}\partial x_{rk}^{l_{rk}}}\right|\end{aligned}$$

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

$$
\begin{aligned}
&\le e^{1/(e-1)}[dA(e-1)]^L \sum_{\sum_{r=1}^J L_r = L} \sum_{(l_{rk})_{r,k}: \sum_{k=0}^K l_{rk} = L_r,\, 1\le r\le J} \prod_{r=1}^J \frac{L_r!}{\prod_{0\le k\le K} l_{rk}!} \\
&= e^{1/(e-1)}[dA(e-1)]^L \sum_{\sum_{r=1}^J L_r = L} \prod_{r=1}^J \left(\sum_{\sum_{0\le k\le K} l_{rk} = L_r} \frac{L_r!}{\prod_{0\le k\le K} l_{rk}!} \right) \\
&= e^{1/(e-1)}[dA(e-1)]^L \sum_{\sum_{r=1}^J L_r = L} \prod_{r=1}^J (K+1)^{L_r} \\
&\le e^{1/(e-1)}[dAJ(e-1)(K+1)]^L,
\end{aligned}
$$

where $d = \|(\xi',X') - (\xi,X)\|$.[22] Note that uniformly for $\|(\xi',X') - (\xi,X)\| \le \frac{1}{2dAJ(e-1)(K+1)}$, the sum of the residuals beyond the Lth term is bounded by $1/2^{L-1}$ and converges to zero as $L \to \infty$. Then, following the proof of Theorem 1, statement (a), we obtain that $\sigma_j(\xi;X,F,\Psi)$ is real analytic with respect to $(\xi,X) \in \mathbb{R}^J \times \mathbb{R}^{J\times K}$ and that its radius of convergence is uniformly bounded away from zero (and it is at least equal to $\frac{1}{2dAJ(e-1)(K+1)}$).

When $G = \Phi$, following the same strategy as for the proof of Theorem 1, statement (b), it suffices to show that κ in Lemma A3, which here equals

$$
\kappa(\tilde{\xi},\tilde{x}) = \int \prod_{j=1}^J \phi(\tilde{\xi}_j + \tilde{x}_j\beta_i) dF(\beta_i)
$$

is real analytic with respect to $(\tilde{\xi},\tilde{x}) \in \mathbb{R}^J \times \mathbb{R}^{J\times K}$. To do so, we control the bound of the higher-order derivative of κ with respect to $(\tilde{\xi},\tilde{x})$. We derive the bound for the case of $\sum_{k=0}^K l_{jk} \ge 1$ for any $j = 1,\ldots,J$, and the bound for the other case, that is, $\sum_{k=0}^K l_{jk} = 0$ for some j, can be obtained in a similar manner. We compute

$$
\begin{aligned}
\frac{\partial^L \kappa(\tilde{\xi},\tilde{x})}{\prod_{j=1}^J \partial \tilde{\xi}_j^{l_{j0}} \prod_{1\le j\le J,\,1\le k\le K} \partial \tilde{x}_{jk}^{l_{jk}}} &= \int \prod_{j=1}^J \frac{\partial^{\sum_{k=0}^K l_{jk}} \phi(\tilde{\xi}_j + \tilde{x}_j\beta_i)}{\partial \tilde{\xi}_j^{l_{j0}} \prod_{k=1}^K \partial \tilde{x}_{jk}^{l_{jk}}} dF(\beta_i) \qquad \textbf{(B.2)} \\
&= \pi^{-\frac{J}{2}}(-1)^{J+L} 2^{-\frac{L}{2}} \int \prod_{j=1}^J \mathrm{H}_{\sum_{k=0}^K l_{jk}-1}\left(\frac{\tilde{\xi}_j + \tilde{x}_j\beta_i}{\sqrt{2}}\right) \exp\left\{-\left(\frac{\tilde{\xi}_j + \tilde{x}_j\beta_i}{\sqrt{2}}\right)^2\right\} \prod_{k=1}^K \beta_{ik}^{\sum_{j=1}^J l_{jk}} dF(\beta_i).
\end{aligned}
$$

Then, using the same techniques as in the proof of Theorem 1, statement (b), and $F \in \mathcal{F}^e$, we obtain

$$
\left| \frac{\partial^L \kappa(\tilde{\xi},\tilde{x})}{\prod_{j=1}^J \partial \tilde{\xi}_j^{l_{j0}} \prod_{1\le j\le J,\,1\le k\le K} \partial \tilde{x}_{jk}^{l_{jk}}} \right| \le A^L (C/\sqrt{2\pi})^J \prod_{j=1}^J \sqrt{(\sum_{k=0}^K l_{jk})!} \qquad \textbf{(B.3)}
$$

[22] The norm $\|\cdot\|$ should be understood as the Euclidean norm on the space of vectorized (ξ,X).

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

and

$$\left|\frac{1}{L!}\left[\sum_{r=1}^{J}(\tilde{\xi}_r'-\tilde{\xi}_r)\frac{\partial}{\partial\tilde{\xi}_r}+(\tilde{x}_{rk}'-\tilde{x}_{rk})\sum_{1\le r\le J,1\le k\le K}\frac{\partial}{\partial\tilde{x}_{rk}}\right]^L\kappa(\tilde{\xi},\tilde{x})\right|$$

$$\le\frac{d^L}{L!}\sum_{\sum_{1\le r\le J,0\le k\le K}l_{rk}=L}\frac{L!}{\prod_{1\le r\le J,0\le k\le K}l_{jk}!}\left|\frac{\partial^L\kappa(\tilde{\xi},\tilde{x})}{\prod_{r=1}^{J}\partial\tilde{\xi}_r^{l_{r0}}\prod_{1\le r\le J,1\le k\le K}\partial\tilde{x}_{rk}^{l_{rk}}}\right|$$

$$\le(C/\sqrt{2\pi})^J[dA]^L\sum_{\sum_{r=1}^J L_r=L}\;\sum_{(l_{rk})_{r,k}:\sum_{k=0}^K l_{rk}=L_r,\,1\le r\le J}\;\prod_{r=1}^{J}\frac{\sqrt{L_r!}}{\prod_{0\le k\le K}l_{rk}!}$$

$$\le(C/\sqrt{2\pi})^J[dA]^L\sum_{\sum_{r=1}^J L_r=L}\frac{1}{\sqrt{\prod_{r=1}^J L_r!}}\prod_{r=1}^{J}\left(\sum_{\sum_{0\le k\le K}l_{rk}=L_r}\frac{L_r!}{\prod_{0\le k\le K}l_{rk}!}\right)$$

$$=(C/\sqrt{2\pi})^J[dA]^L\sum_{\sum_{r=1}^J L_r=L}\frac{\prod_{r=1}^J(K+1)^{L_r}}{\sqrt{\prod_{r=1}^J L_r!}}$$

$$\le(C/\sqrt{2\pi})^J[dA(K+1)]^L\sqrt{\sum_{\sum_{r=1}^J L_r=L}\frac{1}{\prod_{r=1}^J L_r!}\sum_{\sum_{r=1}^J L_r=L}(1)}$$

$$\le(C/\sqrt{2\pi})^J[dA(K+1)J]^L\frac{1}{\sqrt{L!}},$$

where $d=\|(\xi',X')-(\xi,X)\|$. Then, given any d^* and uniformly for $\|(\xi',X')-(\xi,X)\|\le d^*$, the sum of the residuals beyond the Lth term converges to zero as $L\to\infty$. Then, following the proof of Theorem 1, statement (b), we obtain that $\kappa(\tilde{\xi},\tilde{x})$ is real analytic with respect to $(\tilde{\xi},\tilde{x})\in\mathbb{R}^J\times\mathbb{R}^{J\times K}$ with radius of convergence $+\infty$.

Proof of Statement (c). The difference in the proof relative to that of statement (b) in the case of $G=\Phi$ is the bound on $\left|\frac{\partial^L\kappa(\tilde{\xi},\tilde{x})}{\prod_{j=1}^J\partial\tilde{\xi}_j^{l_{j0}}\prod_{1\le j\le J,1\le k\le K}\partial\tilde{x}_{jk}^{l_{jk}}}\right|$, because the moment $\mathbb{E}\left[\prod_{k=1}^K\beta_{ik}^{\sum_{j=1}^J l_{jk}}\right]$ may increase faster when $F\in\mathcal{M}^{f_{e+}}$. Note that

$$\mathbb{E}\left[\prod_{k=1}^K|\beta_{ik}|^{\sum_{j=1}^J l_{jk}}\right]=\sum_{r=1}^m\frac{c_r}{\varsigma_r^K}\int\prod_{k=1}^K|\beta_{ik}|^{\sum_{j=1}^J l_{jk}}f_{e+}\left(\frac{\beta_i-\mu_r}{\varsigma_r}\right)d\beta_i$$

$$=\sum_{r=1}^m c_r\int\prod_{k=1}^K|\varsigma_r t_{ik}+\mu_{rk}|^{\sum_{j=1}^J l_{jk}}f_{e+}(t_i)dt_i$$

$$=\sum_{r=1}^m c_r\varsigma_r^L\int\prod_{k=1}^K\left|t_{ik}+\frac{\mu_{rk}}{\varsigma_r}\right|^{\sum_{j=1}^J l_{jk}}f_{e+}(t_i)dt_i$$

$$\le\sum_{r=1}^m c_r\varsigma_r^L\int\prod_{k=1}^K\sum_{q=0}^{\sum_{j=1}^J l_{jk}}\binom{\sum_{j=1}^J l_{jk}}{q}|t_{ik}|^q\left|\frac{\mu_{rk}}{\varsigma_r}\right|^{\sum_{j=1}^J l_{jk}-q}f_{e+}(t_i)dt_i$$

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

$$
\begin{aligned}
&= \sum_{r=1}^{m} c_r \varsigma_r^L \int \sum_{q_1=0}^{\sum_{j=1}^J l_{j1}} \cdots \sum_{q_K=0}^{\sum_{j=1}^J l_{jK}} \prod_{k=1}^{K} \left[\binom{\sum_{j=1}^J l_{jk}}{q_k} \left| \frac{\mu_{rk}}{\varsigma_r} \right|^{\sum_{j=1}^J l_{jk} - q_k} \right] \prod_{k=1}^{K} |t_{ik}|^{q_k} f_{e+}(t_i) dt_i \\
&\leq \sum_{r=1}^{m} c_r \max\{1,A\}^L \varsigma_r^L \sum_{q_1=0}^{\sum_{j=1}^J l_{j1}} \cdots \sum_{q_K=0}^{\sum_{j=1}^J l_{jK}} \prod_{k=1}^{K} \left[\binom{\sum_{j=1}^J l_{jk}}{q_k} \left| \frac{\mu_{rk}}{\varsigma_r} \right|^{\sum_{j=1}^J l_{jk} - q_k} \right] \prod_{k=1}^{K} \sqrt{q_k!} \\
&\leq \sum_{r=1}^{m} c_r \max\{1,A\}^L \varsigma_r^L \sum_{q_1=0}^{\sum_{j=1}^J l_{j1}} \cdots \sum_{q_K=0}^{\sum_{j=1}^J l_{jK}} \prod_{k=1}^{K} \left[\binom{\sum_{j=1}^J l_{jk}}{q_k} \left| \frac{\mu_{rk}}{\varsigma_r} \right|^{\sum_{j=1}^J l_{jk} - q_k} \right] \prod_{k=1}^{K} \sqrt{\left(\sum_{j=1}^J l_{jk} \right)!} \\
&= \sum_{r=1}^{m} c_r \max\{1,A\}^L \varsigma_r^L \prod_{k=1}^{K} \sqrt{\left(\sum_{j=1}^J l_{jk} \right)!} \prod_{k=1}^{K} \sum_{q=0}^{\sum_{j=1}^J l_{jk}} \binom{\sum_{j=1}^J l_{jk}}{q} \left| \frac{\mu_{rk}}{\varsigma_r} \right|^{\sum_{j=1}^J l_{jk} - q} \\
&\leq [\max\{1,A\}(\bar{\varsigma} + \bar{\mu})]^L \prod_{k=1}^{K} \sqrt{\left(\sum_{j=1}^J l_{jk} \right)!} \\
&\leq \bar{A}^L \prod_{k=0}^{K} \sqrt{\left(\sum_{j=1}^J l_{jk} \right)!},
\end{aligned}
$$

where $\bar{A} = \max\{1,A\}(\bar{\varsigma} + \bar{\mu})$, $\bar{\mu} = \max_{r=1,\ldots,m;\, k=1,\ldots,K}\{|\mu_{rk}|\}$, and $\bar{\varsigma} = \max_{r=1,\ldots,m}\{\varsigma_r\}$. As a result, we only need to multiply the right-hand side of (B.3) by $\bar{A}^L \prod_{k=0}^{K} \sqrt{(\sum_{j=1}^J l_{jk})!}$. Consequently,

$$
\begin{aligned}
&\left| \frac{1}{L!} \left[\sum_{r=1}^{J} (\tilde{\xi}'_r - \tilde{\xi}_r) \frac{\partial}{\partial \tilde{\xi}_r} + (\tilde{x}'_{rk} - \tilde{x}_{rk}) \sum_{1 \leq r \leq J,\, 1 \leq k \leq K} \frac{\partial}{\partial \tilde{x}_{rk}} \right]^L \kappa(\tilde{\xi}, \tilde{x}) \right| \\
&\leq (C/\sqrt{2\pi})^J [d A \bar{A}]^L \sum_{\sum_{r=1}^J L_r = L} \sum_{(l_{rk})_{r,k}: \sum_{k=0}^K l_{rk} = L_r,\, 1 \leq r \leq J} \prod_{r=1}^{J} \frac{L_r!}{\prod_{0 \leq k \leq K} l_{rk}!} \\
&\leq (C/\sqrt{2\pi})^J [d A \bar{A}]^L \sum_{\sum_{r=1}^J L_r = L} \prod_{r=1}^{J} \left(\sum_{\sum_{0 \leq k \leq K} l_{rk} = L_r} \frac{L_r!}{\prod_{0 \leq k \leq K} l_{rk}!} \right) \\
&= (C/\sqrt{2\pi})^J [d A \bar{A}]^L \sum_{\sum_{r=1}^J L_r = L} \prod_{r=1}^{J} (K+1)^{L_r} \\
&\leq (C/\sqrt{2\pi})^J [d A \bar{A} (K+1) J]^L,
\end{aligned}
$$

where $d = \|(\xi', X') - (\xi, X)\|$. Note that uniformly for $\|(\xi', X') - (\xi, X)\| \leq \frac{1}{2A\bar{A}(K+1)J}$, the sum of the residuals beyond the Lth term converges to zero as $L \to \infty$. Then, $\kappa(\tilde{\xi}, \tilde{x})$ is real analytic with respect to $(\tilde{\xi}, \tilde{x}) \in \mathbb{R}^J \times \mathbb{R}^{J \times K}$ and its radius of convergence is uniformly bounded away from zero (and it is at least equal to $\frac{1}{2A\bar{A}(K+1)J}$). The proof is completed.

C. Proofs of Theorem 2 and of Remark 4

Theorem 2, Statement (a). Similar to the proof for the case $\mathcal{M}^{\phi}$ reported in the main text before Theorem 2, the density of the i.i.d. Gumbel distribution satisfies the conditions of Theorem 5(f) in Nguyen et al. (2020). Consequently, the family of finite mixtures $\mathcal{M}^{\psi}$ (and therefore $\mathcal{M}^{\phi}+\mathcal{M}^{\psi}$) has the same approximation property.

Theorem 2, Statement (b). For any distribution F with density f, we construct the following sequence of distributions $\{F_m : m = m_0, m_0+1, \dots\} \subset \mathcal{F}^e$, where $\Pr(\|\beta_i\| \leq m_0) > 0$ and[23]

$$f_m(\beta_i) = \begin{cases} \frac{f(\beta_i)}{\Pr(\|\beta_i\|\leq m)}, & \text{if } \|\beta_i\| \leq m, \\ 0, & \text{otherwise.} \end{cases}$$

Then, as $m \to \infty$, we have

$$\|f_m - f\|_{L_1} = \left(\frac{1}{\Pr(\|\beta_i\| \leq m)} - 1\right)\int_{\|\beta_i\|\leq m} f(\beta_i)d\beta_i + \int_{\|\beta_i\|>m} f(\beta_i)d\beta_i = 2\Pr(\|\beta_i\| > m) \to 0,$$

and therefore

$$\sup_{j,(\xi_t,X_t)} \left|\sigma_j(\xi_t;X_t,F,d_i,G) - \sigma_j(\xi_t;X_t,F_m,d_i,G_m)\right| \tag{C.1}$$

$$\leq \sup_{j,(\xi_t,X_t)} \int \mathbf{1}\{U_{ijt} > U_{ijr}, \forall r \neq j\} \left|f_m(\beta_i;d_i)g_m(\varepsilon) - f(\beta_i;d_i)g(\varepsilon)\right| d\beta_i d\varepsilon$$

$$\leq \sup_{j,(\xi_t,X_t)} \int \mathbf{1}\{U_{ijt} > U_{ijr}, \forall r \neq j\} \left|f_m(\beta_i;d_i) - f(\beta_i;d_i)\right| d\beta_i g_m(\varepsilon)d\varepsilon$$

$$+ \sup_{j,(\xi_t,X_t)} \int \mathbf{1}\{U_{ijt} > U_{ijr}, \forall r \neq j\} \left|g_m(\varepsilon) - g(\varepsilon)\right| d\varepsilon f(\beta_i;d_i)d\beta_i$$

$$\leq \|g_m\|_{L_1}\|f - f_m\|_{L_1} + \|f\|_{L_1}\|g_m - g\|_{L_1}$$

$$\to 0.$$

According to Corollary 1(b), each $\sigma_j(\xi_t;X_t,F_m,d_i,G_m)$ in the sequence is real analytic with respect to $(\xi_t,X_t) \in \mathbb{R}^J \times \mathbb{R}^{J\times K}$.

Theorem 2, Statement (c). Because of the approximation property of Gaussian mixtures, we can find a sequence of distributions $F_m^{\phi} \in \mathcal{M}^{\phi}$ such that $\|f - f_m^{\phi}\|_{L_1} \to 0$ as $m \to \infty$. Similar to the proof of statement (b), we obtain that $\sup_{j,(\xi_t,X_t)} \left|\sigma_j(\xi_t;X_t,F,d_i,G) - \sigma_j(\xi_t;X_t,F_m^{\phi},d_i,G_m^{\phi})\right| \to 0$ with $\sigma_j(\xi_t;X_t,F_m^{\phi},d_i,G_m^{\phi})$ being real analytic with respect to $(\xi_t,X_t) \in \mathbb{R}^J \times \mathbb{R}^{J\times K}$ because of Corollary 1(c) and $\phi \in \mathcal{F}^{e+}$.

Proof of Remark 4. According to Theorem 1 in McFadden and Train (2000, p. 451), any random utility model (RUM) in the class considered by that paper (see McFadden and

[23]The probability $\Pr(\|\beta_i\| \leq m)$ is intended with respect to the distribution F.

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

Train, 2000 for details) can be approximated arbitrarily well by a specification of mixed logit model (2) with linear indices $(\mathcal{X}_j(x_{jt}, d_i)\beta_i)_{j=1}^J$ as in Remark 3. Because we can approximate arbitrarily well any such mixed logit approximant using the models defined in Theorem 2(b) or Theorem 2(c) with the same indices $(\mathcal{X}_j(x_{jt}, d_i)\beta_i)_{j=1}^J$, then we can approximate arbitrarily well the original RUM using these models. Note that these models are real analytic with respect to $(\xi_t, \mathcal{X}_j(x_{jt}, d_i))$ in their domains and $\mathcal{X}_j(x_{jt}, d_i)$ are polynomials of X_t (and therefore real analytic with respect to $X_t \in \mathbb{R}^{J\times K}$). It then follows that these models are real analytic with respect to $(\xi_t, X_t) \in \mathbb{R}^J \times \mathbb{R}^{J\times K}$.

D. Proof of Corollary 3

This proof follows that of Theorem 1 in Wang (2023) and consists of two steps. In the first step, we identify the distribution of $X_t\beta_i$ conditional on X_t. In the second step, using Condition 2 of Corollary 3 and using the same strategy in the proof of Theorem 1 in Wang (2023), we identify the distribution of β_i, F.

We now adapt the proof in Wang (2023) to prove the identification of the distribution of $X_t\beta_i$ conditional on X_t. First, we use Corollary 2(a) to extend the identification of $\sigma(\delta_t; X_t, F, G)$ for $\delta_t \in \Omega_\delta$ to $\mathbb{R}^J$. Second, following the same argument in Appendix A (from equations A.1 to A.4) of Wang (2023) and using Remark 2 in the same appendix, it suffices to prove his Lemma 2 holds when $G \in \mathcal{M}^\phi + \mathcal{M}^\psi$, that is, the zero set of the characteristic function of $\bar{\varepsilon}_i = (\varepsilon_{i1} - \varepsilon_{i0}, \ldots, \varepsilon_{iJ} - \varepsilon_{i0})$ is of zero Lebesgue measure. Following the strategy used in his proof of Lemma 2, it is then sufficient to prove that the real (or imaginary) part of the characteristic function is real analytic. Note that because ε_i is a finite mixture random vector with mixing parameters equal to $(\mu_{i1}, \varsigma_{i1})_{i=1}^{m_1}$ and $(\mu_{i2}, \varsigma_{i2})_{i=1}^{m_2}$ in (B.1), $\bar{\varepsilon}_i$ is then a finite mixture random vector with mixing parameters equal to $(\bar{\mu}_{i1}, \varsigma_{i1})_{i=1}^{m_1}$ and $(\bar{\mu}_{i2}, \varsigma_{i2})_{i=1}^{m_2}$:

$$\bar{G}(\bar{\varepsilon}) = r\sum_{i=1}^{m_1} c_{i1}\frac{1}{\varsigma_{i1}^J}\bar{\Phi}\left(\frac{\bar{\varepsilon} - \bar{\mu}_{i1}}{\varsigma_{i1}}\right) + (1-r)\sum_{i=1}^{m_2} c_{i2}\frac{1}{\varsigma_{i2}^J}\bar{\Psi}\left(\frac{\bar{\varepsilon} - \bar{\mu}_{i2}}{\varsigma_{i2}}\right),$$

where $\bar{\Phi}$ and $\bar{\Psi}$ are the distribution functions of $\bar{\varepsilon}_i$ when ε_i is distributed according to Φ and Ψ with, respectively, $\bar{\mu}_{i1} = (\mu_{ij1} - \mu_{i01})_{j=1}^J$ and $\bar{\mu}_{i2} = (\mu_{ij2} - \mu_{i02})_{j=1}^J$. Then, we can write $\mathbb{E}\left[\exp\{\mathrm{i}t^{\mathrm{T}}\bar{\varepsilon}_i\}\right]$ as

$$\begin{aligned}\mathbb{E}\left[\exp\{\mathrm{i}t^{\mathrm{T}}\bar{\varepsilon}_i\}\right] = r\sum_{i=1}^{m_1} c_{i1}\exp\{\mathrm{i}t^{\mathrm{T}}\bar{\mu}_{i1}\}\mathbb{E}_{\bar{\Phi}}\left[\exp\{\mathrm{i}(\varsigma_{i1}t)^{\mathrm{T}}\bar{\varepsilon}_i\}\right] \\ + (1-r)\sum_{i=1}^{m_2} c_{i2}\exp\{\mathrm{i}\tilde{t}^{\mathrm{T}}\bar{\mu}_{i2}\}\mathbb{E}_{\bar{\Psi}}\left[\exp\{\mathrm{i}(\varsigma_{i2}t)^{\mathrm{T}}\bar{\varepsilon}_i\}\right],\end{aligned}$$

where $\mathbb{E}_{\bar{\Phi}}$ and $\mathbb{E}_{\bar{\Psi}}$ refer to the expectations with respect to $\bar{\Phi}$ and $\bar{\Psi}$, respectively. Note that $\bar{\Phi}$ is still a Gaussian distribution and the real/imaginary part of its characteristic function is real analytic with respect to $t\varsigma_{i1} \in \mathbb{R}^J$ and therefore $t \in \mathbb{R}^J$. Moreover, according to the proof of Lemma 2 in Wang (2023), the real/imaginary part of the characteristic function of $\bar{\Psi}$ is

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

real analytic with respect to $\varsigma_{i2}t \in \mathbb{R}^J$ and therefore $t \in \mathbb{R}^J$. Consequently, the real/imaginary part of $\mathbb{E}\left[\exp\{\mathrm{i}t^{\mathrm{T}}\bar{\varepsilon}_i\}\right]$ is real analytic with respect to $t \in \mathbb{R}^J$. The proof is completed.

E. Proof of Corollary 4

The norm $\|\cdot\|$ of a linear operator $L: E \to W$ is defined as

$$\|L\| := \sup_{v \in E,\, \|v\|_E = 1} \|L(v)\|_W,$$

where $\|\cdot\|_E$ and $\|\cdot\|_W$ denote norms defined in the spaces E and W, respectively. When E and W are Euclidean spaces, we will use the corresponding Euclidean norms. Therefore, $\|L\|$ is the maximal eigenvalue of L.

Using Theorem A in Smale (1986), it suffices to prove that

$$\gamma(\delta) \le \frac{J\|[\partial_\delta \sigma(\delta; X_t, F', G)]^{-1}\| \sup_{k>1}\left[\left(r\left(\frac{C}{\sqrt{2\pi}}\right)^J \frac{1}{\sqrt{k!}} + (1-r)e^{1/(e-1)}(e-1)^k\right)\right]^{\frac{1}{k-1}}}{2\min\{1, \underline{\varsigma}^2\}}.$$

Note that

$$\left\|[\partial_\delta \sigma(\delta; X_t, F', G)]^{-1} \partial_\delta^k \sigma(\delta; X_t, F', G)\right\| \le \|[\partial_\delta \sigma(\delta; X_t, F', G)]^{-1}\| \times \|\partial_\delta^k \sigma(\delta; X_t, F', G)\|,$$

where $\|[\partial_\delta \sigma(\delta; X_t, F', G)]^{-1}\|$ is equal to the reciprocal of the minimal eigenvalue of $\partial_\delta \sigma(\delta; X_t, F', G)$. Moreover, $\partial_\delta^k \sigma(\delta; X_t, F', G) = (\partial_\delta^k \sigma_j(\delta; X_t, F', G))_{j=1}^J$ defines a linear mapping from $\mathbb{R}^{J\times k}$ to $\mathbb{R}^J$ and can be written as: for $m = 1, \dots, J$,

$$\begin{aligned}
\partial_\delta^k \sigma_m(\delta; X_t, F', G)(v_1, \dots, v_k) &= \sum_{(j_1,\dots,j_k): 1 \le j_s \le J,\, s=1,\dots,k} \partial_\delta^k \sigma_m(\delta; X_t, F', G)(e_{j_1}, \dots, e_{j_k}) \prod_{s=1}^k v_{j_s s} \\
&= \sum_{(j_1,\dots,j_k): 1 \le j_s \le J,\, s=1,\dots,k} \frac{\partial^k \sigma_m(\delta; X_t, F', G)}{\prod_{s=1}^k \partial \delta_{j_s}} \prod_{s=1}^k v_{j_s s} \\
&= \sum_{(j_s)_{s=1}^k \in \{1,\dots,J\}^k} \frac{\partial^k \sigma_m(\delta; X_t, F', G)}{\prod_{j=1}^J \partial \delta_j^{\sum_{s=1}^k \mathbf{1}\{j_s = j\}}} \prod_{s=1}^k v_{j_s s},
\end{aligned}$$

where e_j is the standard basis vector with the jth element being 1 and the others being 0 and $\partial_\delta^k \sigma_m(\delta; X_t, F', G)(e_{j_1}, \dots, e_{j_k}) = \frac{\partial^k \sigma_m(\delta; X_t, F', G)}{\prod_{s=1}^k \partial \delta_{j_s}}$. In other words, $\partial_\delta^k \sigma_m(\delta; X_t, F', G)(v_1, \dots, v_k)$ is a polynomial of $(v_1, \dots, v_k)$. Each term in this polynomial corresponds to a term in the polynomial $\prod_{s=1}^k \left(\sum_{j=1}^J v_{js}\right)$. For the term $\prod_{s=1}^s v_{j_s s}$ with $(j_1, \dots, j_k) \in \{1, \dots., J\}^k$, the corresponding coefficient is $\frac{\partial^k \sigma_m(\delta; X_t, F', G)}{\prod_{j=1}^J \partial \delta_j^{\sum_{s=1}^k \mathbf{1}\{j_s = j\}}}$, that is, the $(\sum_{s=1}^k \mathbf{1}\{j_s = 1\}, \dots, \sum_{s=1}^k \mathbf{1}\{j_s = J\})$th derivative of σ_m with respect to δ. Then,

$$\|\partial_\delta^k \sigma_m(\delta;X_t,F',G)\| = \sup_{\|(v_1,\dots,v_k)\|=1} \left|\partial_\delta^k \sigma_m(\delta;X_t,F',G)(v_1,\dots v_k)\right|$$
$$\leq \sup_{\|(v_1,\dots,v_k)\|=1} \left|\frac{\partial_\delta^k \sigma_m(\delta;X_t,F',G)(v_1,\dots v_k)}{\prod_{s=1}^k \left(\sum_{j=1}^J |v_{js}|\right)}\right| \times \sup_{\|(v_1,\dots,v_k)\|=1} \prod_{s=1}^k \left(\sum_{j=1}^J |v_{js}|\right)$$
$$\leq \max_{(l_1,\dots,l_J):\sum l_j=k} \left|\frac{\partial^k \sigma_m(\delta;X_t,F',G)}{\prod_{j=1}^J \partial\delta_j^{l_j}}\right| \times \sqrt{\frac{J^k}{k^k}}.$$

Using Lemmas A2 and A3, we then obtain

$$\|\partial_\delta^k \sigma_m(\delta;X_t,F',G)\| \leq \frac{1}{\underline{\varsigma}^k}\left(r\left(\frac{C}{\sqrt{2\pi}}\right)^J \sqrt{k!} + (1-r)e^{1/(e-1)}(e-1)^k k!\right)\sqrt{\frac{J^k}{k^k}}$$
$$= \frac{k!}{\underline{\varsigma}^k}\left(r\left(\frac{C}{\sqrt{2\pi}}\right)^J \frac{1}{\sqrt{k!}} + (1-r)e^{1/(e-1)}(e-1)^k\right)\sqrt{\frac{J^k}{k^k}}.$$

Because $\|[\partial_\delta \sigma(\delta;X_t,F',G)]^{-1}\| > 1$ and $C \geq \sqrt{2\pi}$, we finally obtain

$$\gamma(\delta) \leq \sup_{k>1}\left[\|[\partial_\delta \sigma(\delta;X_t,F',G)]^{-1}\frac{1}{\underline{\varsigma}^k}\left(r\left(\frac{C}{\sqrt{2\pi}}\right)^J \frac{1}{\sqrt{k!}} + (1-r)e^{1/(e-1)}(e-1)^k\right)\sqrt{\frac{J^k}{k^k}}\right]^{\frac{1}{k-1}}$$
$$\leq \|[\partial_\delta \sigma(\delta;X_t,F',G)]^{-1}\| \frac{J\sup_{k>1}\left[\left(r\left(\frac{C}{\sqrt{2\pi}}\right)^J \frac{1}{\sqrt{k!}} + (1-r)e^{1/(e-1)}(e-1)^k\right)\right]^{\frac{1}{k-1}}}{2\min\{1,\underline{\varsigma}^2\}}.$$

E.1. The Case of the Mixed Logit Model

In the case of the mixed logit model, $r = 0$ and $\underline{\varsigma} = 1$ in the expression of $\gamma(\delta)$. Then,

$$\gamma(\delta) \leq \|[\partial_\delta \sigma(\delta;X_t,F',G)]^{-1}\| \frac{Je^{1/(e-1)}(e-1)^2}{2} \approx 2.64J \times \|[\partial_\delta \sigma(\delta;X_t,F',G)]^{-1}\|.$$

We now bound the eigenvalues of $\partial_\delta \sigma(\delta;X_t,F',G)$. Note that the Jacobian matrix

$$\partial_\delta \sigma(\delta;X_t,F',G) = \int \left[\sigma_{ij}\mathbf{1}\{j=r\} - \sigma_{ij}\sigma_{ir}\right]_{j,r} dF(\beta_i) = \int \left[\mathrm{Diag}(\sigma_i) - \sigma_i\sigma_i^{\mathrm{T}}\right] dF(\beta_i),$$

where $\sigma_i = (\sigma_{i1},\dots\sigma_{iJ})$ denotes the vector of multinomial choice probabilities with the random coefficients β equal to β_i. Denote by λ_i the minimal eigenvalue of $\mathrm{Diag}(\sigma_i) - \sigma_i\sigma_i^{\mathrm{T}}$ and v_i the corresponding eigenvector. Without loss of generality, suppose that the maximal element of v_i in absolute value is its first coordinate $v_{i1} \neq 0$. Then, for any $v \in \mathbb{R}^J$,

$$v^{\mathrm{T}}\partial_\delta \sigma(\delta;X_t,F',G)v = \int v^{\mathrm{T}}\left[\mathrm{Diag}(\sigma_i) - \sigma_i\sigma_i^{\mathrm{T}}\right] v dF(\beta_i)$$
$$\implies v^{\mathrm{T}}\partial_\delta \sigma(\delta;X_t,F',G)v \geq \int \lambda_i \|v\|^2 dF(\beta_i).$$

Moreover, we have

$$\begin{aligned}\left[\text{Diag}(\sigma_i)-\sigma_i\sigma_i^{\mathrm{T}}\right]v_i=\lambda_i v_i &\Longrightarrow \sigma_{i1}(v_{i1}-\sum_{j=1}^{J}\sigma_{ij}v_{ij})=\lambda_i v_{i1}\\ &\Longrightarrow \lambda_i=\sigma_{i1}\left(1-\frac{\sum_{j=1}^{J}\sigma_{ij}v_{ij}}{v_{i1}}\right)\\ &\geq \sigma_{i1}\left(1-\frac{\sum_{j=1}^{J}\sigma_{ij}|v_{ij}|}{|v_{i1}|}\right)\\ &\geq \sigma_{i1}\left(1-\sum_{j=1}^{J}\sigma_{ij}\right)\\ &\geq \sigma_{i1}\sigma_{i0},\end{aligned}$$

where $\sigma_{i0}=1-\sum_{j=1}^{J}\sigma_{ij}$. Then, any eigenvalue of $\partial_\delta\sigma(\delta;X_t,F',G)$ is greater than or equal to $\int\sigma_{i0}\min_{1\leq j\leq J}\sigma_{ij}dF(\beta_i)\approx\sigma_0\min_{1\leq j\leq J}\sigma_j$.[24] Then, plugging this inequality into the expression for Δ^*, we obtain the desired approximation of Δ^*.

To obtain Δ_{H}^* in (18), note that, in the classic BLP fixed point algorithm based on contraction mappings, $\delta^{(n+1)}-\delta^{(n)}$ is

$$\begin{aligned}\delta^{(n+1)}-\delta^{(n)}&=\ln\sigma(\delta^{(n)};X_t,F',G)-\ln p_t\\ &\approx \text{Diag}(1/p_{1t},\ldots,1/p_{Jt})(\sigma(\delta^{(n)};X_t,F',G)-p_t)\\ &=-\text{Diag}(1/p_{1t},\ldots,1/p_{Jt})\left[\partial_\delta\sigma(\delta^{(N)};X_t,F',G)\right]\left(\delta_{\mathrm{NR}}^{(n+1)}-\delta^{(n)}\right),\end{aligned}$$

where $\delta_{\mathrm{NR}}^{(n+1)}-\delta^{(n)}$ refers to the $(n+1)$th step in NR algorithm (14) if one switches from the FP step to the NR step at $\delta^{(n)}$. Then,

$$\left\|\delta^{(n+1)}-\delta^{(n)}\right\|\approx\left\|\text{Diag}(1/p_{1t},\ldots,1/p_{Jt})\left[\partial_\delta\sigma(\delta^{(N)};X_t,F',G)\right]\left(\delta_{\mathrm{NR}}^{(n+1)}-\delta^{(n)}\right)\right\|.$$

When $\delta^{(N)}$ is close to the unique solution, the minimal eigenvalue of

$$\text{Diag}(1/p_{1t},\ldots,1/p_{Jt})\left[\partial_\delta\sigma(\delta^{(N)};X_t,F',G)\right]$$

is approximately p_{0t}. Then, $\left\|\delta_{\mathrm{NR}}^{(n+1)}-\delta^{(n)}\right\|$ is approximately bounded by $\left\|\delta_{\mathrm{NR}}^{(n+1)}-\delta^{(n)}\right\|/p_{0t}$, and $\left\|\delta_{\mathrm{NR}}^{(N+1)}-\delta^{(N)}\right\|<\Delta_{\mathrm{H}}^*=\Delta^*p_{0t}$ implies $\left\|\delta_{\mathrm{NR}}^{(N+1)}-\delta^{(N)}\right\|<\Delta^*$, that is, $\delta^{(N)}$ is an approximate zero.

[24] More precisely, $\text{Cov}(\sigma_{i0},\min_{1\leq j\leq J}\sigma_{ij})=\int\sigma_{i0}\min_{1\leq j\leq J}\sigma_{ij}dF(\beta_i)-\sigma_0\min_{1\leq j\leq J}\sigma_j$. The approximation holds when $\text{Cov}(\sigma_{i0},\min_{1\leq j\leq J}\sigma_{ij})$ is small.

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

F. Proof of Corollary 5

LEMMA F1. *Suppose $\tau(\theta):\Theta\to\mathbb{R}$ is a positive real analytic function in $\Theta\subset\mathbb{R}^m$ and $\tau(\theta)\geq a$ for any $\theta\in\Theta$. Moreover, for any $(l_1,\dots,l_m)$ with $\sum_{k=1}^m l_k=L$, we have*

$$\left|\frac{\partial^L\tau(\theta)}{\prod_{k=1}^m\partial\theta_k^{l_k}}\right|\leq AA_\tau^L\prod_{k=1}^m l_k!, \tag{F.1}$$

where A and A_τ are constants that do not depend on θ. Then, for $L\geq 1$,

$$\left|\frac{\partial^L\ln\tau(\theta)}{\prod_{k=1}^m\partial\theta_k^{l_k}}\right|\leq\left(\frac{A_\tau(a+1)^{1/m}}{(a+1)^{1/m}-1}\right)^L\prod_{k=1}^m l_k!. \tag{F.2}$$

Proof. Without loss of generality, suppose $A=1$. Otherwise, we can normalize $\tau(\theta)$ to $\tau(\theta)/A$ and any derivative of $\ln(\tau(\theta)/A)$ will be equal to that of $\ln\tau(\theta)$.

The techniques used in this proof are similar to those used in Lemmas A1 and A2. Denote $B=\frac{(a+1)^{1/m}}{(a+1)^{1/m}-1}$. By induction, we first prove the lemma for $L=1$. For $k=1,\dots,m$, we have

$$\left|\frac{\partial\ln\tau(\theta)}{\partial\theta_k}\right|=\left|\frac{\frac{\partial\tau(\theta)}{\partial\theta_k}}{\tau(\theta)}\right|\leq\frac{A_\tau}{a}<\frac{A_\tau(a+1)^{1/m}}{(a+1)^{1/m}-1}.$$

The lemma then holds for $L=1$. Now suppose that the lemma holds up to $L-1$. Then, we have

$$\begin{aligned}
&\frac{\partial\ln\tau(\theta)}{\partial\theta_k}=\frac{\frac{\partial\tau(\theta)}{\partial\theta_k}}{\tau(\theta)}\\
&\Longrightarrow\tau(\theta)\frac{\partial\ln\tau(\theta)}{\partial\theta_k}=\frac{\partial\tau(\theta)}{\partial\theta_k}\\
&\Longrightarrow\sum_{r_1=0}^{l_1-1}\sum_{r_2=0}^{l_2}\cdots\sum_{r_m=0}^{l_m}\binom{l_1-1}{r_1}\binom{l_2}{r_2}\cdots\binom{l_m}{r_m}\frac{\partial^{\sum_{k=1}^m(l_k-r_k)}\ln\tau(\theta)}{\prod_{k=1}^m\theta_k^{l_k-r_k}}\frac{\partial^{\sum_{k=1}^m r_k}\tau(\theta)}{\prod_{k=1}^m\theta_k^{r_k}}=\frac{\partial^L\tau(\theta)}{\prod_{k=1}^m\theta_k^{l_k}}\\
&\Longrightarrow\left|\frac{\partial^L\ln\tau(\theta)}{\prod_{k=1}^m\partial\theta_k^{l_k}}\right|\\
&\quad\leq\frac{1}{a}\left(\sum_{0\leq r_1\leq l_1-1,0\leq r_k\leq l_k\forall k\geq 2,\sum r_k>0}\binom{l_1-1}{r_1}\prod_{k=2}^m\binom{l_k}{r_k}\left|\frac{\partial^{\sum_{k=1}^m(l_k-r_k)}\ln\tau(\theta)}{\prod_{k=1}^m\theta_k^{l_k-r_k}}\frac{\partial^{\sum_{k=1}^m r_k}\tau(\theta)}{\prod_{k=1}^m\theta_k^{r_k}}\right|\right.\\
&\qquad\left.+\left|\frac{\partial^L\tau(\theta)}{\prod_{k=1}^m\theta_k^{l_k}}\right|\right)\\
&\Longrightarrow\frac{1}{A_\tau^L\prod l_k!}\left|\frac{\partial^L\ln\tau(\theta)}{\prod_{k=1}^m\partial\theta_k^{l_k}}\right|\leq\frac{1}{a}\left(\sum_{0\leq r_1\leq l_1-1,0\leq r_k\leq l_k\forall k\geq 2,\sum r_k>0}\frac{l_1-r_1}{l_1}\frac{1}{A_\tau^{\sum(l_k-r_k)}\prod(l_k-r_k)!}\right.\\
&\quad\left.\left|\frac{\partial^{\sum_{k=1}^m(l_k-r_k)}\ln\tau(\theta)}{\prod_{k=1}^m\theta_k^{l_k-r_k}}\right|\frac{1}{A_\tau^{\sum r_k}\prod r_k!}\left|\frac{\partial^{\sum_{k=1}^m r_k}\tau(\theta)}{\prod_{k=1}^m\theta_k^{r_k}}\right|+\frac{1}{A_\tau^L\prod l_k!}\left|\frac{\partial^L\tau(\theta)}{\prod_{k=1}^m\theta_k^{l_k}}\right|\right)
\end{aligned}$$

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

$$\Longrightarrow \frac{1}{A_\tau^L \prod l_k!}\left|\frac{\partial^L \ln \tau(\theta)}{\prod_{k=1}^m \partial\theta_k^{l_k}}\right| \leq \frac{1}{a}\left(1+\sum_{r_1=0}^{l_1-1}\sum_{r_2=0}^{l_2}\cdots\sum_{r_m=0}^{l_m} B^{\sum_{k=1}^m (l_k-r_k)} - B^L\right)$$

$$\leq \frac{1}{a}\left(1+B^L\sum_{r_1=1}^{l_1-1} B^{-r_1}\sum_{r_2=0}^{l_2} B^{-r_2}\cdots\sum_{r_m=0}^{l_m} B^{-r_m} - B^L\right)$$

$$\Longrightarrow \frac{1}{A_\tau^L \prod l_k!}\left|\frac{\partial^L \ln \tau(\theta)}{\prod_{k=1}^m \partial\theta_k^{l_k}}\right| \leq \frac{1}{a}\left(1+B^L\frac{1-B^{-l_1}}{1-B^{-1}}\prod_{k=2}^m \frac{B^{-l_k-1}}{1-B^{-1}} - B^L\right)$$

$$\leq \frac{1}{a}\left(1+B^L\frac{1}{(1-B^{-1})^m} - B^L\right) = B^L.$$

Consequently, the lemma holds for any $(l_1,\dots,l_m)$ with $\sum l_k = L$. The proof is completed. □

We now prove Corollary 5 for the case of $G \in \mathcal{M}^\phi + \mathcal{M}^\psi$ and $F_v \in \mathcal{F}^e$ using Lemma F1. Given y_{it}, the likelihood function $\ell(\xi,\Sigma;x_{it},y_{ij}) = \ln\int \sigma_{iy_{it}}(\xi_t + x_{it}\Sigma v_i;G)dF_v(v_i)$. First, for any $j=1,\dots,J$, applying Corollary 1(b), $\int \sigma_{ij}(\xi_t + x_{it}\Sigma v_i;G)dF_v(v_i)$ is real analytic with respect to $(\xi_t, x_{it}\Sigma)$ in their domains and its higher-order derivatives satisfy (F.1) with the corresponding constant A being equal to $r(C/\sqrt{2\pi})^J + (1-r)e^{1/(e-1)}$ and A_τ being equal to $\frac{A_v(e-1)}{\underline{\varsigma}}$, where C is a constant defined in (A.11). Then, $\int \sigma_{ij}(\xi_t + x_{it}\Sigma v_i;G)dF_v(v_i)$ is real analytic with respect to (ξ_t,Σ) in their domains and its higher-order derivatives satisfy (F.1) with A_τ equal to $\frac{A_v J(e-1)|x_{it}|_{\max}}{\underline{\varsigma}}$. Denote $m = JT+KP$. Using Lemma F1, we obtain that the higher-order derivatives of ℓ_{it} are bounded by $\left(\frac{A_v J(e-1)\underline{\varsigma}^{-1}|x_{it}|_{\max}(a+1)^{1/m}}{(a+1)^{1/m}-1}\right)^L \prod_{k=1}^m l_k!$. Consequently, the higher-order derivatives of $\mathcal{L}(\xi,\Sigma)$ are bounded by $\left(\frac{A_v J(e-1)\underline{\varsigma}^{-1}(a+1)^{1/m}}{(a+1)^{1/m}-1}\right)^L \prod_{k=1}^m l_k!\sum_{i,t}|x_{it}|_{\max}^L$. Following the arguments in the proof of Corollary 4, we then obtain

$$\sup_{\|(v_1,\dots,v_k)\|=1,\, v_1,\dots,v_k\in\mathbb{R}^m} \|\partial^k_{(\xi,\,\Sigma)}\mathcal{L}(\xi,\Sigma)(v_1,\dots,v_k)\|$$

$$\leq \left(\frac{A_v J(e-1)\underline{\varsigma}^{-1}(a+1)^{1/m}}{(a+1)^{1/m}-1}\right)^k k!\sqrt{\frac{m^k}{k^k}}\sum_{i,t}|x_{it}|_{\max}^k$$

and

$$\gamma(\xi,\Sigma) \leq \sup_{k>1}\left\|\left[\frac{\partial^2_{(\xi,\,\Sigma)}\mathscr{L}(\xi,\Sigma)}{NT}\right]^{-1}\right\|^{1/(k-1)}$$

$$\times\left[\left(\frac{A_v J(e-1)\underline{\varsigma}^{-1}(a+1)^{1/m}}{(a+1)^{1/m}-1}\right)^{k+1}\sqrt{\frac{m^{k+1}}{(k+1)^{k+1}}}\frac{\sum_{i,t}|x_{it}|_{\max}^{k+1}}{NT}\right]^{1/(k-1)}$$

$$\leq \max\left\{1, \left\|\left[\frac{\partial^2_{(\xi,\,\Sigma)}\mathscr{L}(\xi,\Sigma)}{NT}\right]^{-1}\right\|\right\}\left(\max\left\{1, \frac{A_v J(e-1)\underline{\varsigma}^{-1}\bar{x}(a+1)^{1/m}}{(a+1)^{1/m}-1}\right\}\right)^3\left(\frac{m}{3}\right)^{3/2}.$$

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

To prove the second statement, it suffices to replace $\gamma(\xi,\Sigma)$ in the definition of $\alpha(\xi,\Sigma)$ by the upper bound above. The proof is completed.

G. Monte Carlo Simulations: Details

In this Appendix, we detail the data generating processes and the fixed point algorithm used in Section 6. Our Monte Carlo setting resembles that in Section 5 of Conlon and Gortmaker (2020). Because our Monte Carlo experiments focus on the numerical performance of different implementations of demand inverses rather than the GMM estimator, we simulate data for only one market and ignore the market index t.

There are two firms that sell products $j = 1,\ldots,\lfloor J/2\rfloor$ and $j = \lfloor J/2\rfloor+1,\ldots,J$, respectively, with constant marginal costs of production $c_j = [1,x_j,w_j]\gamma+\omega_j$. We specify

$$U_{ij} = \beta_{i0}\text{price}_j + \beta_{ix}x_j + \alpha + \xi_j + \varepsilon_{ij},$$

where $(\varepsilon_{ij})_{j=0}^J$ is i.i.d. Gumbel (i.e., which gives rise to a mixed logit) and $(\text{price}_j)_{j=1}^J$ is generated by a simultaneous Bertrand price-setting game as in Conlon and Gortmaker (2020). Demand and supply shocks (ξ_j,ω_j) follow a mean-zero Gaussian distribution with $\sigma_\xi^2 = \sigma_\omega^2 = 0.2$ and $\sigma_{\xi\omega} = 0.1$, and are independent across $j = 1,\ldots,J$. Moreover, $\alpha = -1$ and (β_{i0},β_{ix}) follows a Gaussian distribution:

$$(\beta_{i0},\beta_{ix}) \sim \mathcal{N}\left((-2,6),\begin{bmatrix}\sigma_p^2 & 0\\ 0 & 9\end{bmatrix}\right).$$

The configuration "No random coefficient on price" in Tables 1 and 2 refers to $\sigma_p = 0$, that is, $\beta_{i0} = -2$ for any i, and "Random coefficient on price" refers to $\sigma_p = 0.2$. To simulate market shares and prices, we use 1,000 independent draws of (β_{i0},β_{ix}) and for $J = 25,50,100$, where $J = 50$ is equal to the maximal number of products in the configurations considered by Conlon and Gortmaker (2020).

The fixed point algorithm we use in Section 6 and in the FP step of Algorithm 1 is the original one by Berry et al. (1995): for $j = 1,\ldots,J$,

$$\delta_j^{(n+1)} = \delta_j^{(n)} - (\ln p_j - \ln\sigma_j(\delta^{(n)};\text{price},x,G)). \qquad \textbf{(G.1)}$$

"Close starting values" in Table 1 refer to a situation where we draw starting values from a small neighborhood of the true values $(\delta_j)_{j=1}^J = (-2\text{price}_j + 6x_j - 1 + \xi_j)_{j=1}^J$. In practice, we draw 200 starting values from the uniform distribution on a "tight" neighborhood of the true values $[\delta_1-1,\delta_1+1]\times\cdots\times[\delta_J-1,\delta_J+1]$. Instead, "Distant starting values" in Table 2 refer to a situation where we draw the same number of starting values from a uniform distribution on the "wider" neighborhood of the true values $[\delta_1-5,\delta_1+5]\times\cdots\times[\delta_J-5,\delta_J+5]$.

REFERENCES

Aguirregabiria, V., & Mira, P. (2007). Sequential estimation of dynamic discrete games. *Econometrica*, 75(1), 1–53.

Allen, R., & Rehbeck, J. (2020). Identification of random coefficient latent utility models. Available at SSRN 3545696.

https://doi.org/10.1017/S0266466624000148 Published online by Cambridge University Press

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

Berry, S., Levinsohn, J., & Pakes, A. (2004). 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. T., & Haile, P. A. (2014). Identification in differentiated products markets using market level data. *Econometrica*, 82(5), 1749–1797.

Berry, S. T., & Haile, P. A. (2018). Identification of nonparametric simultaneous equations models with a residual index structure. *Econometrica*, 86(1), 289–315.

Compiani, G. (2022). Market counterfactuals and the specification of multiproduct demand: A nonparametric approach. *Quantitative Economics*, 13(2), 545–591.

Conlon, C., & Gortmaker, J. (2020). Best practices for differentiated products demand estimation with PyBLP. *RAND Journal of Economics*, 51(4), 1108–1161.

Dubé, J.-P., Fox, J. T., & Su, C.-L. (2012a). Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica*, 80(5), 2231–2267.

Dubois, P., Griffith, R., & O'Connell, M. (2020). How well targeted are soda taxes? *American Economic Review*, 110(11), 3661–3704.

Fox, J. T., & Gandhi, A. (2016). Nonparametric identification and estimation of random coefficients in multinomial choice models. *RAND Journal of Economics*, 47(1), 118–139.

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

Goolsbee, A., & Petrin, A. (2004). The consumer gains from direct broadcast satellites and the competition with cable TV. *Econometrica*, 72(2), 351–381.

Gowrisankaran, G., & Rysman, M. (2012). Dynamics of consumer demand for new durable goods. *Journal of Political Economy*, 120(6), 1173–1219.

Hausman, J. A., & Wise, D. A. (1978). A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica*, 46, 403–426.

Iskhakov, F., Lee, J., Rust, J., Schjerning, B., & Seo, K. (2016). Comment on "constrained optimization approaches to estimation of structural models". *Econometrica*, 84(1), 365–370.

Knittel, C. R., & Metaxoglou, K. (2014). Estimation of random-coefficient demand models: Two empiricists' perspective. *Review of Economics and Statistics*, 96(1), 34–59.

Krasikov, I. (2004). New bounds on the Hermite polynomials. *East Journal on Approximations*, 10(3), 355–362.

Lang, S. (2012). *Real and functional analysis* (Vol. 142). Springer Science & Business Media.

Lee, J., & Seo, K. (2016). Revisiting the nested fixed-point algorithm in BLP random coefficients demand estimation. *Economics Letters*, 149, 67–70.

Masten, M. A. (2018). Random coefficients on endogenous variables in simultaneous equations models. *Review of Economic Studies*, 85(2), 1193–1250.

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

Mugnier, M., & Wang, A. (2022). Identification and (fast) estimation of large nonlinear panel models with two-way fixed effects. Available at SSRN 4186349.

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

Nguyen, T. T., Nguyen, H. D., Chamroukhi, F., & McLachlan, G. J. (2020). Approximation by finite mixtures of continuous density functions that vanish at infinity. *Cogent Mathematics & Statistics*, 7(1), 1750861.

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy*, 110(4), 705–729.

Rheinboldt, W. C. (1988). On a theorem of S. Smale about Newton's method for analytic mappings. *Applied Mathematics Letters*, 1(1), 69–72.

Rudin, W. (1976). *Principles of mathematical analysis*. (3rd ed.) McGraw-Hill.

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, 55, 999–1033.

Salanié, B., & Wolak, F. A. (2019). Fast, "robust," and approximately correct: Estimating mixed demand systems. Discussion paper, National Bureau of Economic Research.

Seim, K. (2006). An empirical model of firm entry with endogenous product-type choices. *RAND Journal of Economics*, 37(3), 619–640.

Smale, S. (1986). Newton's method estimates from data at one point. In *The merging of disciplines: New directions in pure, applied, and computational mathematics: Proceedings of a Symposium Held in Honor of Gail S. Young at the University of Wyoming, August 8–10, 1985. Sponsored by the Sloan Foundation, the National Science Foundation, and Air Force Office of Scientific Research* (pp. 185–196). Springer.

Smith, H. (1935). Discontinuous demand curves and monopolistic competition: A special case. *Quarterly Journal of Economics*, 49(3), 542–550.

Stinchcombe, M. B., & White, H. (1998). Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory*, 14(3), 295–325.

Su, C.-L., & Judd, K. L. (2012). Constrained optimization approaches to estimation of structural models. *Econometrica*, 80(5), 2213–2230.

Train, K. E., & Winston, C. (2007). Vehicle choice behavior and the declining market share of US automakers. *International Economic Review*, 48(4), 1469–1496.

Wang, A. (2023). Sieve BLP: A semi-nonparametric model of demand for differentiated products. *Journal of Econometrics*, 235(2), 325–351.