

Submitted to

Kuhn–Tucker Demand Models: Scalable Estimation and Universal Approximation

Xiao Lei

HKU Business School, The University of Hong Kong, Hong Kong SAR xlei@hku.hk

Authors are encouraged to submit new papers to INFORMS journals by means of a style file template, which includes the journal title. However, use of a template does not certify that the paper has been accepted for publication in the named journal. INFORMS journal templates are for the exclusive purpose of submitting to an INFORMS journal and are not intended to be a true representation of the article's final published form. Use of this template to distribute papers in print or online or to submit papers to another non-INFORM publication is prohibited.

Abstract. Kuhn–Tucker (KT) demand models characterize consumers' budget allocation through the Karush–Kuhn–Tucker conditions of utility maximization. Classical specifications, including quadratic utility, are widely used to study pricing, assortment, and competition, but remain difficult to estimate from consumer data. By contrast, models designed for tractable estimation often impose restrictive assumptions on demand behavior. We address this gap with two results. First, for KT models whose utilities are linear in their coefficients, we establish convexity of maximum-likelihood estimation under Gaussian shocks through an equivalent chance-constrained reformulation. The same reformulation reduces likelihood evaluation to at most one-dimensional integration, regardless of the number of goods, enabling scalable estimation with global optimality guarantees. Second, we show that mixtures of these estimable KT models can approximate a broad class of consumption distributions to arbitrary accuracy across varying budgets, prices, and assortments, including behavior generated by nonconcave utilities. Numerical studies demonstrate estimation with 50 goods, compared with at most 10 goods for conventional methods within the same time limit, and improved empirical fit from richer utility specifications.

Key words: multiple discrete–continuous choice; demand estimation; chance-constrained optimization; mixture models; universal approximation

1. Introduction

Economic and business research frequently studies how consumers allocate a limited budget across multiple goods. A standard approach assumes that consumers maximize a concave utility function subject to a budget constraint. Let q_j denote the quantity consumed of good $j = 1, \dots, J$, $p_j > 0$ its price, and $m > 0$ the consumer's budget. Writing $q = (q_1, \dots, q_J)^\top$, the consumer solves

$$\max_{q \geq 0} v(q) \quad \text{subject to} \quad \sum_{j=1}^J p_j q_j \leq m.$$

For a differentiable concave utility, the Karush–Kuhn–Tucker (KKT) conditions characterize the optimal allocation, including which goods are not consumed. Models based on this characterization are known as

Kuhn–Tucker (KT) demand models (Wales and Woodland 1983; Bhat 2008). A classical example is concave quadratic utility, which is widely used to study pricing, assortment, and competition because of its simple structure (Singh and Vives 1984; Rochet and Choné 1998; Federgruen and Hu 2015).

Estimating these models from consumer data is much less straightforward than using them for decision analysis. To account for unobserved preferences, we write utility as $v(q, \epsilon)$, where ϵ is a vector of preference shocks observed by the consumer but not by the researcher. This stochastic version of the choice model is called a multiple discrete–continuous (MDC) choice model, which describes which goods consumers purchase and in what quantities. The MDC literature develops econometric models and estimation methods for these choices, together with their empirical applications. Surprisingly, quadratic utility and other concave specifications widely used in decision analysis have seen limited use in MDC applications, even though Wales and Woodland (1983) developed a quadratic MDC model more than four decades ago. Traditional likelihood formulations require multidimensional integration over unobserved shocks, repeated across observations and parameter updates. Furthermore, previous studies typically fit KT models using quasi-Newton methods without establishing the convexity of the maximum likelihood estimation (MLE) problem. Numerical convergence therefore does not guarantee a global optimum. Due to the estimation challenge, the most popular specification in the MDC literature is the multiple discrete–continuous extreme value (MDCEV) model, whose translated-power utility, combined with independent extreme-value shocks, yields a concave log-likelihood with a closed-form expression (Bhat 2008). However, this utility form is less convenient for analyzing pricing and competition. There is therefore a clear gap: models useful for decision analysis remain difficult to estimate at scale, while models that are readily estimable are less convenient for decision analysis. This gap raises an important unresolved question: *can we develop scalable estimation methods with global optimality guarantees for general KT models, so that classical models of economic decision making can also be estimated reliably from consumer data?*

The structure that makes MDCEV convenient to estimate also restricts the demand behavior it can represent. Its additively separable utility cannot capture direct complementarities between goods. Its independent extreme-value shocks impose substitution restrictions analogous to the independence of irrelevant alternatives (IIA) property in the multinomial logit (MNL) model (Garikapati et al. 2014). In discrete choice modeling, a standard way to expand the range of behavior a model can represent is to use mixtures of models with different preference types. For example, mixtures of MNL models can approximate a broad class of random-utility choice probabilities (McFadden and Train 2000). However, as we show later, mixtures of MDCEV models have fundamental limitations and cannot explain even simple consumption patterns. This motivates a second research question: *can mixtures of a reasonable class of utility models approximate a broad class of consumer behavior, even when the underlying utilities are nonconcave?*

We address these questions for KT models whose utilities are linear in the coefficients to be estimated. This class includes quadratic utility, as well as commonly used translated logarithmic (Phaneuf et al.

2000) and translated-power (Bhat 2008) utility specifications. For the first research question, we prove that maximum-likelihood estimation under Gaussian shocks is a convex optimization problem in the utility coefficients, so that standard quasi-Newton methods can converge to a global optimum. The key step is to reformulate maximum-likelihood estimation as an equivalent chance-constrained optimization (CCO) problem. The likelihood combines the density of purchased quantities with the conditional probability that unpurchased-good shocks ϵ are low enough to make zero consumption optimal. The difficulty is that these shock restrictions involve the unobserved shadow price of the budget constraint. Through a change of variables, we eliminate this shadow price and express the restrictions as linear inequalities involving a set of independent Gaussian variables. Maximizing the likelihood can then be formulated using chance constraints that require these inequalities to hold jointly with sufficiently high probability. Standard convexity results for chance-constrained optimization establish the convexity of this formulation (Shapiro, Dentcheva, and Ruszczyński 2009). Furthermore, the same reformulation reduces likelihood evaluation to at most one-dimensional integration, regardless of the number of goods. This substantially reduces the computational burden of estimation. The convexity result extends to all log-concave shock distributions. It also holds (1) when we add an outside option with concave utility; (2) when we allow consumers to face multiple resource constraints, such as limits on both time and money; and (3) when consumers pay an entry fee in addition to usage charges or decide whether to purchase an indivisible good alongside their choices of continuous quantities. With participation fees or indivisible goods, the consumer’s decision problem can be nonconvex, but estimation remains convex.

For the second research question, our goal is to approximate a target model’s consumption distribution uniformly across varying budgets, prices, and assortments of available goods in the range of interest. While MDCEV has been regarded as a counterpart of MNL in the MDC literature, we first show that standard mixed MDCEV lacks the approximation flexibility of mixed MNL. For example, it cannot capture households buying fewer instant noodles on average as their food budgets increase. We then show that mixtures of the KT models covered by our convexity result can uniformly approximate the joint consumption distributions generated by a broad class of continuous utilities, including nonconcave utilities. Our estimation and approximation results together yield flexible demand mixtures that can be estimated using expectation-maximization (EM) with convex M-steps. Therefore, mixtures of KT models can approximate complex consumption behavior to arbitrary accuracy while retaining an interpretable economic model of each consumer type, whose optimal decisions are characterized by the KKT conditions.

In our numerical experiments, the proposed estimator scales to demand systems with 50 goods and accommodates nonlinear utilities, whereas conventional likelihood methods and optimal transport complete estimation only for systems with at most 10 goods within the same time limit. Across two empirical applications, a KT specification combining quadratic and logarithmic utility achieves better in-sample and

out-of-sample fit than MDCEV. The improvement persists when both models are extended to mixtures of consumer types.

In summary, we establish MDC counterparts of two foundational results in discrete choice: convex and scalable estimation of MNL (McFadden 1974) and universal approximation by mixed MNL (McFadden and Train 2000). By connecting maximum-likelihood estimation with chance-constrained optimization, we also offer a new perspective on demand estimation and provide a new econometric application of chance-constrained optimization.

2. Literature Review

Models of Multiproduct Consumption. Economic and business research uses models of multiproduct consumption to understand demand and analyze firms’ decisions. Over the past decade, these models have attracted growing attention in the operations community. This literature follows two broad modeling approaches. The first draws on the economic tradition of modeling consumption through utility maximization over continuous quantities (Singh and Vives 1984, Rochet and Choné 1998). In operations management, such models have been used to study pricing (Huh and Li 2022), assortment (Zhang et al. 2023), competition (Caro and Martínez-de Albéniz 2012, Federgruen and Hu 2015), and content rotation (Lei et al. 2023). The second builds on discrete-choice modeling, extending logit (Tulabandhula et al. 2023, Bai et al. 2024, Jasin et al. 2024), sequential-choice (Ferreira and Goh 2021, Ke and Wang 2022, El Housni et al. 2025), ranking-based (Lin et al. 2026), and other random-utility (Abdallah et al. 2026) or perturbed-utility (Liu et al. 2025) models to accommodate consumers purchasing multiple products. We contribute to the first approach by developing a general estimation framework for utility-based models of continuous consumption, enabling researchers to support theoretical insights by calibrating these models to consumer data.

Choice Model Estimation and Representation. Our work addresses two longstanding topics in choice modeling: scalable estimation and flexible representations of consumer behavior. On estimation, computational approaches include expectation–maximization algorithms (Şimşek and Topaloglu 2018) and column generation for discovering preference types (van Ryzin and Vulcano 2015). Another approach exploits the optimization structure of estimation problems. Jagabathula, Subramanian, and Venkataraman (2020) use a convex formulation of mixing-distribution estimation to develop a conditional-gradient algorithm that generates mixture components iteratively. Related work exploits the likelihood structure of tree logit models (Jagabathula et al. 2024) and uses inverse optimization to learn correlated preferences (Sun 2026). On representation, mixed MNL can approximate a broad class of random utility models (McFadden and Train 2000). Ranking-based models (Farias et al. 2013), Markov-chain models (Blanchet et al. 2016), decision forests (Chen and Mišić 2022), market segmentation trees (Aouad et al. 2023), and neural networks (Aouad and Désir 2026) provide alternative ways to represent consumer preferences and choice behavior. Further work characterizes relationships among choice-model classes (Feng et al. 2017). Most of this literature

studies discrete choice models. We extend scalable convex estimation and flexible mixture approximation to MDC models, where consumers jointly choose products and consumption quantities. Although other approaches, such as neural networks, may also achieve universal approximation in the MDC setting, KT mixtures combine this flexibility with an explicit utility-maximization model for each consumer type.

MDC Model Estimation. Estimating KT demand models has long been computationally challenging because zero consumption gives rise to multivariate probability integrals (Lee and Pitt 1986). For Gaussian specifications, these integrals are typically evaluated using deterministic approximations (Dutt 1976, Wales and Woodland 1983) or simulation (Kim et al. 2002, Millimet and Tchernis 2008). Even with these methods, the reported applications and numerical experiments involve at most ten goods. An alternative is to choose utility and shock specifications jointly to simplify the likelihood. The popular MDCEV model combines a particular translated-power utility with extreme-value shocks, yielding a closed-form likelihood (Bhat 2008). For general KT models, however, computational difficulties remain a barrier to empirical adoption (Pinjari et al. 2024). Our approach enables estimation of general KT demand models at a substantially larger scale, with numerical experiments reaching 50 goods compared with at most ten in the studies above.

Chance-Constrained Optimization. Chance-constrained optimization has been applied to inventory control (Jiang et al. 2019), call-center staffing (Gurvich et al. 2010), project planning (Shen et al. 2010), portfolio selection (Bonami and Lejeune 2009), and power-system operations (Bienstock et al. 2014). Our work extends the scope of chance-constrained optimization to econometric demand estimation.

3. Estimating KT Models

We begin with a KT demand model whose systematic utility is strongly concave in quantities and linear in the coefficients to be estimated. Section 3.2 uses quadratic utility to explain the likelihood and its estimation difficulties. Section 3.3 reformulates that likelihood and establishes log-likelihood concavity, and Section 3.4 describes computation and identification. Section 3.5 considers outside consumption, multiple resource constraints, participation fees, and binary choices.

3.1. The Baseline MDC Model

There are J goods. A context $c = (p, m, A)$ consists of positive unit resource requirements $p \in \mathbb{R}_{++}^J$, a resource budget $m > 0$, and a nonempty assortment $A \subseteq \{1, \dots, J\}$ of available goods. When the resource is money, p_j is the price of good j and m is the expenditure budget. When the resource is time, p_j is the time required per unit of activity j and m is the available time. The feasible allocations are

$$K(c) = \{q \in \mathbb{R}_+^J : p^\top q \leq m, q_j = 0 \text{ for } j \notin A\}. \quad (1)$$

A consumer privately observes a vector of utility shocks $\epsilon \in \mathbb{R}^J$ and chooses

$$q_\theta(c, \epsilon) = \arg \max_{q \in K(c)} \{v_\theta(q) + \epsilon^\top q\}. \quad (2)$$

Here v_θ is the systematic utility, indexed by an unknown parameter vector $\theta \in \Theta \subseteq \mathbb{R}^L$, where Θ is a convex parameter set. We assume that v_θ is twice continuously differentiable on a neighborhood of the feasible quantity domains and uniformly strongly concave in q for every $\theta \in \Theta$. Specifically, let $H_\theta(q) = -\nabla_q^2 v_\theta(q)$ denote the negative Hessian of utility with respect to quantities. There exists a constant $\mu > 0$ such that $H_\theta(q) \succeq \mu I_J$ for every $\theta \in \Theta$ and every feasible q , where I_J is the J -dimensional identity matrix. The consumer's problem in (2) therefore has a unique solution, characterized by its Karush–Kuhn–Tucker (KKT) optimality conditions.

Within this framework, we focus on systematic utilities that are linear in their unknown parameters $\theta = (\theta_1, \dots, \theta_L)$. Specifically,

$$v_\theta(q) = \sum_{\ell=1}^L \theta_\ell \psi_\ell(q), \quad \theta \in \Theta, \quad (3)$$

where $\psi_1, \dots, \psi_L$ are known, twice continuously differentiable basis functions. The basis functions may include linear terms q_j , quadratic terms $q_j q_k$, and other nonlinear terms such as $\log(1 + q_j)$. Linear-in-parameters utility specifications are standard in choice and demand modeling (Train 2009; Satomura, Kim, and Allenby 2011).

The shocks ϵ have a known joint probability distribution that is independent of the context c . For the remainder of the paper, we focus on standard Gaussian shocks, $\epsilon \sim N(0, I_J)$. Gaussian errors are commonly used in the choice-modeling literature, where the resulting specifications are referred to as probit models (e.g., Train 2009). For an assortment A , let f_A denote the joint density of the shocks $(\epsilon_j)_{j \in A}$.

Our baseline accommodates classical utility specifications, including quadratic utility, which serves as the running example below. Translated-power and logarithmic forms used in Bhat (2008) also fit the fixed-basis specification for v_θ when their shape parameters are fixed and the strong-concavity condition holds. The baseline model assumes either that there is no outside option or that the outside option provides zero marginal utility. Omitting an outside good is a standard specification in the MDC literature (Kim, Allenby, and Rossi 2002; Bhat 2008). Section 3.5.1 relaxes this assumption by allowing unused resources to contribute to utility.

3.2. Estimation Challenge

The data consist of n independent observations (q_i, p_i, m_i, A_i) , $i = 1, \dots, n$; the utility shocks ϵ_i are unobserved. Write $c_i = (p_i, m_i, A_i)$. Treating budgets as observed is common in MDC estimation (Wales and Woodland 1983; Kim, Allenby, and Rossi 2002; Bhat 2022). Let $f_\theta(q_i | c_i)$ denote the joint density of q_i conditional on c_i . We estimate the utility coefficients by maximizing the log likelihood

$$\max_{\theta \in \Theta} \ell_n(\theta), \quad \ell_n(\theta) = \sum_{i=1}^n \log f_\theta(q_i | c_i).$$

To illustrate the estimation procedure in the literature, consider the quadratic–Gaussian model of (2), where $v_{a,B}(q) = a^\top q - \frac{1}{2} q^\top B q$ and $B = B^\top \succeq \mu I_J$. Here $a \in \mathbb{R}^J$ gives systematic marginal utilities at zero

consumption, and $B \in \mathbb{R}^{J \times J}$ describes how marginal utilities change with own consumption and the consumption of other goods. Estimation requires learning a and all distinct entries of B .

Let $S_i = \{j \in A_i : q_{ij} > 0\}$ denote the purchased goods. The observed quantities q_i are optimal if and only if there exists a budget multiplier λ_i satisfying the following KKT conditions:

$$\begin{aligned} \epsilon_{ij} &= (Bq_i - a)_j + \lambda_i p_{ij}, \quad j \in S_i, & (\text{KKT-a}) \\ \epsilon_{ij} &\leq (Bq_i - a)_j + \lambda_i p_{ij}, \quad j \in A_i \setminus S_i, & (\text{KKT-b}) \\ q_i &\geq 0, \quad p_i^\top q_i \leq m_i, & (\text{KKT-c}) \\ q_{ij} &= 0, \quad j \notin A_i, & (\text{KKT-d}) \\ \lambda_i &\geq 0, & (\text{KKT-e}) \\ \lambda_i (m_i - p_i^\top q_i) &= 0. & (\text{KKT-f}) \end{aligned}$$

When the budget is not exhausted, (KKT-f) gives $\lambda_i = 0$. For a purchased good $j \in S_i$, marginal utility must be zero at the optimal quantity, so (KKT-a) fixes its shock at $\epsilon_{ij} = (Bq_i - a)_j$. For an unpurchased good $j \in A_i \setminus S_i$, marginal utility at zero consumption must be nonpositive, so (KKT-b) requires $\epsilon_{ij} \leq (Bq_i - a)_j$. Consequently,

$$\begin{aligned} f_{a,B}(q_i | c_i) &= \text{joint density of } q_i \\ &= C_i(B) \times \text{joint density of } (\epsilon_{ij})_{j \in S_i} \times \Pr\{\epsilon_{ij} \leq (Bq_i - a)_j, \quad j \in A_i \setminus S_i \mid \epsilon_{ij} = (Bq_i - a)_j, \quad j \in S_i\} \\ &= C_i(B) \prod_{j \in S_i} \phi((Bq_i - a)_j) \times \Pr\{\epsilon_{ij} \leq (Bq_i - a)_j, \quad j \in A_i \setminus S_i \mid \epsilon_{ij} = (Bq_i - a)_j, \quad j \in S_i\} \\ &= C_i(B) \prod_{j \in S_i} \phi((Bq_i - a)_j) \prod_{j \in A_i \setminus S_i} \Phi((Bq_i - a)_j). \end{aligned}$$

Here, $C_i(B)$ is the Jacobian factor that converts the density of the purchased goods' shocks into the density of q_i . For the fixed observation, this is a positive scalar depending on B . The remaining equalities use independence of the standard Gaussian shocks, whose density and cumulative distribution function are denoted by ϕ and Φ .

When the budget is exhausted, we may have $\lambda_i > 0$. A common approach in the literature removes λ_i from the shock restrictions by taking differences between the normalized shocks ϵ_{ij}/p_{ij} . Choose a purchased good $r_i \in S_i$ and define

$$\eta_{ir_i} = \frac{\epsilon_{ir_i}}{p_{ir_i}}, \quad \eta_{ij} = \frac{\epsilon_{ij}}{p_{ij}} - \frac{\epsilon_{ir_i}}{p_{ir_i}}, \quad j \in A_i \setminus \{r_i\}.$$

These transformed shocks are jointly Gaussian. Although the original shocks ϵ_{ij} are independent, the differences η_{ij} are correlated because they share the reference shock ϵ_{ir_i} . Conditions (KKT-a), (KKT-b), and (KKT-e) give, respectively,

$$\begin{aligned} \eta_{ij} &= \frac{(Bq_i - a)_j}{p_{ij}} - \frac{(Bq_i - a)_{r_i}}{p_{ir_i}}, \quad j \in S_i \setminus \{r_i\}, \\ \eta_{ij} &\leq \frac{(Bq_i - a)_j}{p_{ij}} - \frac{(Bq_i - a)_{r_i}}{p_{ir_i}}, \quad j \in A_i \setminus S_i, \\ \eta_{ir_i} &\geq \frac{(Bq_i - a)_{r_i}}{p_{ir_i}}. \end{aligned}$$

The first line fixes the difference between the normalized shocks of each purchased good $j \in S_i \setminus \{r_i\}$ and the reference good r_i . The second line requires the normalized shocks of unpurchased goods to be sufficiently small relative to the reference shock. The third line is equivalent to $\lambda_i \geq 0$. Thus, the purchased-good shock differences contribute their joint density, and we integrate over the remaining shocks subject to the other two restrictions. That is,

$$\begin{aligned} f_{a,B}(q_i | c_i) &= \text{joint density of } q_i \\ &= C_i^\Delta(B) \times \text{joint density of } (\eta_{ij})_{j \in S_i \setminus \{r_i\}} \\ &\quad \times \Pr \left\{ \begin{array}{l} \eta_{ij} \leq \frac{(Bq_i - a)_j}{p_{ij}} - \frac{(Bq_i - a)_{r_i}}{p_{ir_i}}, j \in A_i \setminus S_i, \\ \eta_{ir_i} \geq \frac{(Bq_i - a)_{r_i}}{p_{ir_i}} \end{array} \middle| \begin{array}{l} \eta_{ij} = \frac{(Bq_i - a)_j}{p_{ij}} - \frac{(Bq_i - a)_{r_i}}{p_{ir_i}}, \\ j \in S_i \setminus \{r_i\} \end{array} \right\}. \end{aligned} \quad (4)$$

Here, $C_i^\Delta(B)$ is the Jacobian factor for this change of variables. The conditional probability is a multivariate normal integral over the transformed shocks of the unpurchased goods and the reference good r_i . Wales and Woodland (1983) compute the likelihood using approximations to multivariate normal integrals. For a similar MDC model with independent normal log-taste shocks, Kim, Allenby, and Rossi (2002) use the Geweke–Hajivassiliou–Keane (GHK) simulator to evaluate the likelihood within a hierarchical Bayesian estimator.

This approach presents several estimation challenges. First, multivariate probabilities must be evaluated for each observation and each trial parameter vector. Direct integration becomes costly as the number of goods J increases, while simulation and analytical approximations introduce error into the likelihood used for estimation. Second, a smooth approximation to the likelihood function permits numerical optimization, including quasi-Newton methods, but smoothness alone does not guarantee joint concavity in the utility coefficients or global optimality. For quadratic utility, another challenge is to ensure that B remains positive definite during estimation. Wales and Woodland (1983) do not specify how this constraint is enforced during optimization. Lewbel and Nesheim (2019) use a factorized quadratic specification to ensure utility concavity, but the estimation problem is nonconvex under their parameterization.

3.3. Chance-Constrained Estimation and Log-Likelihood Concavity

We now present an alternative way to eliminate the budget multiplier. This approach expresses the likelihood as a deterministic factor multiplied by the probability that independent standard Gaussian variables satisfy a set of inequalities, leading to a chance-constrained formulation of maximum likelihood.

We first develop the formulation using the quadratic–Gaussian model with an exhausted budget. For an observation i with purchased set S_i , we need to record only $|S_i| - 1$ purchased quantities. Choosing $r_i \in S_i$, the remaining quantity is $q_{ir_i} = (m_i - \sum_{j \in S_i \setminus \{r_i\}} p_{ij} q_{ij}) / p_{ir_i}$. The purchased-good conditions in (KKT-a) give $\epsilon_{ij} = (Bq_i - a)_j + \lambda_i p_{ij}$ for $j \in S_i$. For fixed coefficients a, B , these equations provide an affine change

of variables between the $|S_i|$ purchased shocks and the $|S_i|$ variables consisting of the $|S_i| - 1$ purchased quantities and λ_i . The quantities and the multiplier are jointly determined by the original shocks. We can therefore obtain their joint density from the purchased-shock density and the Jacobian of this transformation. The unpurchased shocks must also satisfy (KKT-b). Integrating each of them up to its threshold in (KKT-b) gives

$$\begin{aligned} g_{a,B}(q_i, \lambda \mid c_i) &= C_i^\lambda(B) \prod_{j \in S_i} \phi((Bq_i - a)_j + \lambda p_{ij}) \times \prod_{j \in A_i \setminus S_i} \left[\int_{-\infty}^{(Bq_i - a)_j + \lambda p_{ij}} \phi(u) du \right] \\ &= C_i^\lambda(B) \prod_{j \in S_i} \phi((Bq_i - a)_j + \lambda p_{ij}) \prod_{j \in A_i \setminus S_i} \Phi((Bq_i - a)_j + \lambda p_{ij}), \quad \lambda \geq 0. \end{aligned} \quad (5)$$

Here, $g_{a,B}$ is written in terms of $|S_i| - 1$ purchased quantities and λ_i , instead of all $|S_i|$ purchased quantities. The factor $C_i^\lambda(B)$ is the Jacobian for this variable transformation. Note that the data record the quantities but not the multiplier. Different multiplier values correspond to different shocks that can support the same observed allocation. To obtain the density of the quantities alone, we therefore integrate the joint density over all feasible multiplier values:

$$\begin{aligned} f_{a,B}(q_i \mid c_i) &= \int_0^\infty g_{a,B}(q_i, \lambda \mid c_i) d\lambda \\ &= C_i^\lambda(B) \int_0^\infty \prod_{j \in S_i} \phi((Bq_i - a)_j + \lambda p_{ij}) \prod_{j \in A_i \setminus S_i} \Phi((Bq_i - a)_j + \lambda p_{ij}) d\lambda. \end{aligned} \quad (6)$$

To express the integral in (6) as a Gaussian probability, define

$$\bar{\lambda}_i(a, B) = -\frac{\sum_{j \in S_i} p_{ij}(Bq_i - a)_j}{\sum_{j \in S_i} p_{ij}^2}, \quad \sigma_i^2 = \frac{1}{\sum_{j \in S_i} p_{ij}^2}, \quad (7)$$

and

$$D_i(a, B) = C_i^\lambda(B) \sigma_i (2\pi)^{-(|S_i|-1)/2} \exp \left[-\frac{1}{2} \left(\sum_{j \in S_i} (Bq_i - a)_j^2 - \frac{\bar{\lambda}_i(a, B)^2}{\sigma_i^2} \right) \right]. \quad (8)$$

The likelihood in (6) can then be reformulated as

$$f_{a,B}(q_i \mid c_i) = D_i(a, B) \int_0^\infty \frac{1}{\sigma_i} \phi \left(\frac{\lambda - \bar{\lambda}_i(a, B)}{\sigma_i} \right) \prod_{j \in A_i \setminus S_i} \Phi((Bq_i - a)_j + \lambda p_{ij}) d\lambda.$$

For any given a, B , the factor $D_i(a, B)$ is fully determined by the observation. The remaining integral is now taken against the $N(\bar{\lambda}_i(a, B), \sigma_i^2)$ density in λ . To express this integral using a fixed standard Gaussian distribution, substitute $\lambda = \bar{\lambda}_i(a, B) + \sigma_i z$. The differential $d\lambda = \sigma_i dz$ cancels the density factor $1/\sigma_i$, and the lower bound becomes $z = -\bar{\lambda}_i/\sigma_i$. Thus,

$$f_{a,B}(q_i \mid c_i) = D_i(a, B) \int_{-\bar{\lambda}_i/\sigma_i}^\infty \phi(z) \prod_{j \in A_i \setminus S_i} \Phi((Bq_i - a)_j + p_{ij}(\bar{\lambda}_i + \sigma_i z)) dz. \quad (9)$$

Let $Z_i \sim N(0, 1)$, independently of the unpurchased-good shocks. Conditional on $Z_i = z$, the product of Gaussian distribution functions in (9) is the probability that all unpurchased-good inequalities hold. Integrating this conditional probability against $\phi(z)$, over the specified range, gives $f_{a,B}(q_i | c_i) = D_i(a, B) \times P_i(a, B)$, where

$$P_i(a, B) = \Pr \left\{ \begin{array}{l} \bar{\lambda}_i(a, B) + \sigma_i Z_i \geq 0, \\ \epsilon_{ij} \leq (Bq_i - a)_j + p_{ij}(\bar{\lambda}_i(a, B) + \sigma_i Z_i), \quad j \in A_i \setminus S_i \end{array} \right\}. \quad (10)$$

The first inequality preserves the nonnegative-multiplier restriction. The remaining inequalities preserve the unpurchased-good KKT conditions. When the budget is not exhausted, $\lambda_i = 0$, and Section 3.2 gives the same density–probability decomposition with

$$\begin{aligned} D_i(a, B) &= C_i(B) \prod_{j \in S_i} \phi((Bq_i - a)_j), \\ P_i(a, B) &= \Pr \{ \epsilon_{ij} \leq (Bq_i - a)_j, \quad j \in A_i \setminus S_i \} \\ &= \prod_{j \in A_i \setminus S_i} \Phi((Bq_i - a)_j), \end{aligned} \quad (11)$$

where $C_i(B)$ is the corresponding Jacobian factor. This case requires only univariate Gaussian probabilities.

For either budget case, taking logarithms gives $\log f_{a,B}(q_i | c_i) = \log D_i(a, B) + \log P_i(a, B)$. The first term has a closed-form expression in a, B . To express the second term through an optimization constraint, introduce s_i satisfying $s_i \leq \log P_i(a, B) \iff P_i(a, B) \geq e^{s_i}$. This has the standard joint chance-constraint form used in the CCO literature (Nemirovski and Shapiro 2006): the inequalities defining $P_i(a, B)$ must hold jointly with probability at least e^{s_i} . For a sample containing either type of observation, maximum likelihood can consequently be written as

$$\begin{aligned} \max_{(a,B) \in \Theta, s \in \mathbb{R}^n} \quad & \sum_{i=1}^n [\log D_i(a, B) + s_i] \\ \text{subject to} \quad & P_i(a, B) \geq e^{s_i}, \quad i = 1, \dots, n, \end{aligned} \quad (12)$$

where D_i and P_i use the appropriate expressions for each observation's budget status. For fixed a, B , increasing s_i increases the objective. Maximization therefore sets s_i to its largest feasible value, $\log P_i(a, B)$. Substituting these values recovers the complete sample log likelihood, establishing equivalence between the chance-constrained formulation and maximum likelihood.

To establish that (12) is a convex optimization problem, we show that its feasible set is convex and its objective is concave. First consider the chance constraints. Rearranging the inequalities in (10) gives

$$P_i(a, B) = \Pr \left\{ \begin{array}{l} -\sigma_i Z_i \leq \bar{\lambda}_i(a, B), \\ \epsilon_{ij} - p_{ij}\sigma_i Z_i \leq (Bq_i - a)_j + p_{ij}\bar{\lambda}_i(a, B), \quad j \in A_i \setminus S_i \end{array} \right\}.$$

The random variables on the left have a fixed joint Gaussian distribution. The thresholds on the right are affine in a, B : this follows directly from the definition of $\bar{\lambda}_i(a, B)$ in (7). By Shapiro, Dentcheva, and

Ruszczynski (2009, Theorem 4.39), $P_i(a, B)$ is log-concave in a, B , so $\log P_i(a, B)$ is concave. For an unexhausted budget, the thresholds are simply $(Bq_i - a)_j$, so the same argument applies. Consequently, $s_i - \log P_i(a, B)$ is convex jointly in s_i, a, B , and the chance constraint $s_i - \log P_i(a, B) \leq 0$ defines a convex region. Together with the convex parameter domain Θ , these constraints give a convex feasible set for (12).

We next examine the objective. Since each s_i enters linearly, it remains to establish concavity of $\log D_i(a, B)$. For an exhausted budget, (7)–(8) give

$$\log D_i(a, B) = \log C_i^\lambda(B) - \frac{1}{2} \sum_{j \in S_i} ((Bq_i - a)_j + p_{ij} \bar{\lambda}_i(a, B))^2 + \log \sigma_i - \frac{|S_i| - 1}{2} \log(2\pi).$$

Each expression inside a square is affine in a, B , so its negative square is concave. The last two terms are constant with respect to the coefficients. We therefore need to establish concavity of the remaining term, $\log C_i^\lambda(B)$. Let $H_i(B)$ denote the negative Hessian of utility with respect to the $|S_i| - 1$ purchased quantities, after substituting the budget equality for the remaining quantity. Its positive definiteness follows from the strong concavity of the original utility. The Jacobian factor satisfies

$$C_i^\lambda(B) = \underbrace{\text{positive factor determined by prices}}_{\text{constant with respect to } B} \times \det H_i(B).$$

Since $H_i(B)$ is affine in B and positive definite, the concavity of the log determinant implies that $\log C_i^\lambda(B)$ is concave in B . When the budget is not exhausted, $C_i(B) = \det(B_{S_i S_i})$, so its logarithm is also concave in B . Thus, $\log D_i(a, B)$ is concave for both budget cases. Together with the convex chance constraints, this establishes that (12) is a convex optimization problem. For the general utilities in Section 3.1, replace $(Bq_i - a)_j$ in the shock expressions by $-\nabla v_\theta(q_i)_j$, which is affine in θ under (3). The same argument then establishes log-concavity of the Gaussian density and probability factors. The relevant negative Hessian remains positive definite by strong concavity and affine in the coefficients under (3). Thus, the log-Jacobian term is also concave, and the same convexity result holds.

THEOREM 1 (Convex estimation). *The maximum-likelihood estimation problem for the models in Section 3.1 is a convex optimization problem.*

The result also extends from independent Gaussian shocks to any given log-concave shock distribution. In the general likelihood formula, the integrand is jointly log-concave in the coefficients and the variables being integrated out. Integration preserves log-concavity (Shapiro, Dentcheva, and Ruszczyński 2009, Theorems 4.15 and 4.26), while the Jacobian retains the log-determinant structure above. Appendix A.1 provides the complete proof.

3.4. Estimation and Identification

Our concavity result in Theorem 1 allows standard constrained quasi-Newton methods to converge to a global optimum of $\max_{\theta \in \Theta} \ell_n(\theta)$ under standard convergence conditions. The same reformulation also resolves the multidimensional integration difficulty described in Section 3.2. For an exhausted budget, (6) involves a single integral over λ , whose integrand is a product of univariate Gaussian densities and cumulative distribution functions. For an unexhausted budget, (11) is evaluated directly using ϕ and Φ . The gradient with respect to the utility coefficients follows by differentiating these expressions. Thus, evaluating the likelihood and its gradient requires at most one-dimensional numerical integration, regardless of the number of goods. The same calculation applies to the general utilities in Section 3.1.

For quadratic utility, we enforce positive definiteness of B through the convex constraint $B \succeq \mu I_J$. At each iteration, we minimize a quadratic approximation to the negative log likelihood subject to this constraint to obtain a search direction. This subproblem can be written as a semidefinite program (SDP); a line search then determines the step size. Section 5 gives the SDP formulation used in the numerical experiments.

We next consider population identification: whether the conditional distribution of quantities determines systematic utility up to an additive constant. Such a constant does not affect consumption choices.

PROPOSITION 1 (Population identification). *At a fixed context c , two systematic utilities satisfying the assumptions in Section 3.1 generate the same conditional distribution of quantities if and only if they differ by a constant on $K(c)$.*

The proof builds on the uniqueness argument in Carlier, Chernozhukov, and Galichon (2016): two utilities generating the same quantity distribution must prescribe the same quantities for the same shocks. Consider any feasible allocation at which all available goods are purchased and the budget is not exhausted. Choosing the shocks to offset the systematic marginal utilities makes this allocation optimal under the first utility, and hence under the second. The first-order conditions then imply that the two utilities have the same marginal utilities at this allocation. The utilities therefore differ by a constant over all such allocations, and continuity extends this conclusion to all of $K(c)$. Appendix A.2 provides the proof.

3.5. Practical Considerations

We next examine whether estimation remains convex when the baseline model incorporates additional features of consumer decisions.

3.5.1. Outside Consumption Unspent money or time can be used for other goods or activities. MDC models have represented the value of these outside options using linear utility (Palma and Hess 2022), logarithmic utility (Phaneuf, Kling, and Herriges 2000), and power utility (Kim, Allenby, and Rossi 2002; Bhat 2008). Let $y \geq 0$ denote the amount of resources devoted to outside consumption and $h_\eta(y)$ its utility. As in (3), we specify $h_\eta(y) = \sum_r \eta_r \chi_r(y)$, where the known basis functions χ_r are twice continuously

differentiable for $y > 0$. We restrict η to a convex set on which h_η is strictly increasing and concave, and define $h_\eta(0)$ by its right limit. Since outside utility is strictly increasing, the consumer allocates the entire budget between the goods in A and outside consumption:

$$\max_{q, y \geq 0} \{v_\theta(q) + h_\eta(y) + \epsilon^\top q\}, \quad p^\top q + y = m, \quad q_j = 0 \text{ for } j \notin A.$$

We observe the total budget m , so outside consumption is determined by $y = m - p^\top q$. The coefficients θ and η are estimated jointly.

COROLLARY 1 (Estimation with outside consumption). *Under the assumptions in Section 3.1 and the outside-utility specification above, maximum-likelihood estimation is a convex optimization problem in the utility coefficients (θ, η) .*

To see why, we substitute $y = m - p^\top q$ into the objective. This gives the choice problem in (2), with systematic utility

$$\tilde{v}_{\theta, \eta}(q; c) = v_\theta(q) + h_\eta(m - p^\top q). \quad (13)$$

This utility is linear in (θ, η) , and adding the concave outside term preserves strong concavity in q . For $y > 0$, it also satisfies the smoothness requirement in Section 3.1, so Theorem 1 applies to the combined coefficient vector.

3.5.2. Multiple Resource Constraints Leisure activities require both money and time, while grocery purchases may be limited by money and storage space. Exhausting one resource can prevent further consumption even when another remains available. Satomura, Kim, and Allenby (2011) and Mondal and Bhat (2021) develop likelihood formulations and estimation procedures for such models, but do not establish joint concavity of the log likelihood in the estimated parameters.

To incorporate these constraints, let D be the matrix of observed resource requirements, with one row per resource and one column per available good, and let b be the vector of observed resource limits. The constraints are $q \geq 0$ and $Dq \leq b$. We assume that the feasible set is nonempty and compact and retain the utility and shock assumptions of Section 3.1.

COROLLARY 2 (Estimation with multiple resource constraints). *Under the assumptions in Section 3.1 and the resource-constraint specification above, maximum-likelihood estimation is a convex optimization problem in the utility coefficients θ .*

For an observed allocation q_i , let R_i denote the binding resource constraints and $\lambda_{ik} \geq 0$ the multiplier for resource $k \in R_i$. The purchased- and unpurchased-good conditions become

$$\begin{aligned} \epsilon_{ij} &= -[\nabla v_\theta(q_i)]_j + \sum_{k \in R_i} D_{i,kj} \lambda_{ik}, \quad j \in S_i, \\ \epsilon_{ij} &\leq -[\nabla v_\theta(q_i)]_j + \sum_{k \in R_i} D_{i,kj} \lambda_{ik}, \quad j \in A_i \setminus S_i. \end{aligned}$$

If the $|R_i|$ binding constraints are linearly independent when restricted to the purchased goods, they determine $|R_i|$ purchased quantities from the remaining $|S_i| - |R_i|$ quantities. As in Section 3.3, we can therefore transform the $|S_i|$ purchased-good shocks into $|S_i| - |R_i|$ purchased quantities and $|R_i|$ multipliers. Integrating over the multipliers and unpurchased-good shocks subject to the optimality conditions gives the quantity likelihood. The shock restrictions remain affine in the utility coefficients and the variables being integrated out, while the Jacobian retains the log-determinant structure in Section 3.3. The same argument therefore establishes log-likelihood concavity. The resulting likelihood and its gradient can be evaluated using at most a $|R_i|$ -dimensional integral, with one integration variable for each independent binding resource constraint.

3.5.3. Participation Fees and Binary Choices Fixed participation fees and indivisible goods add discrete decisions to the consumer’s allocation problem. We show that maximum-likelihood estimation remains convex even when the consumer’s problem is nonconvex.

Participation fees. Consuming a good may require paying a fixed cost in addition to its usage cost. Car use, for example, involves fixed ownership costs and expenses that increase with driving distance, as modeled by Tanner and Bolduc (2012). The same structure arises with a time budget: exercising at a gym requires travel and preparation time in addition to the time spent exercising. Let $F_j \geq 0$ denote a known fixed cost incurred whenever $q_j > 0$. We retain the baseline utility and replace its budget constraint with $p^\top q + \sum_{j \in A} F_j \mathbf{1}\{q_j > 0\} \leq m$. Choosing among purchased sets can make the overall problem nonconvex.

Binary choices. Consumers may also choose indivisible goods together with continuous quantities. A classic example is the joint choice of appliance ownership and electricity consumption (Dubin and McFadden 1984). Under a time budget, a person may decide whether to attend a scheduled concert and how much of the remaining evening to spend reading or exercising. Let $x_k \in \{0, 1\}$, $k = 1, \dots, K$, indicate whether binary option k is selected, and let ρ_k denote its known monetary or time cost. Systematic utility $v_\theta(q, x)$ may include interactions such as $q_j x_k$, which capture how appliance ownership changes the value of electricity consumption. Each binary option has a standard Gaussian utility shock ξ_k , independent across options and of ϵ . The consumer solves

$$\begin{aligned} \max_{q \geq 0, x \in \{0, 1\}^K} \quad & v_\theta(q, x) + \epsilon^\top q + \xi^\top x \\ \text{subject to} \quad & p^\top q + \rho^\top x \leq m, \quad q_j = 0 \text{ for } j \notin A. \end{aligned}$$

As in Section 3.1, we specify $v_\theta(q, x) = \sum_{\ell=1}^L \theta_\ell \psi_\ell(q, x)$, with known functions ψ_ℓ and a convex coefficient domain. For every fixed x , utility is smooth and strongly concave in q .

COROLLARY 3 (Estimation with participation fees and binary choices). *Under the assumptions in Section 3.1 and the participation-fee and binary-choice specifications above, maximum-likelihood estimation is a convex optimization problem in the utility coefficients θ .*

We explain the result using the binary-choice model. With the observed x_i fixed, the consumer faces a continuous choice problem with remaining budget $m_i - \rho_i^\top x_i$, so the change of variables in Section 3.3 applies. The observed choice (q_i, x_i) must also yield at least as much utility as every feasible alternative (q, x) . Equivalently,

$$\epsilon_i^\top (q - q_i) + \xi_i^\top (x - x_i) \leq v_\theta(q_i, x_i) - v_\theta(q, x) = \sum_{\ell=1}^L \theta_\ell [\psi_\ell(q_i, x_i) - \psi_\ell(q, x)].$$

Adding these linear constraints to the inequalities in Section 3.3 preserves convexity. The Gaussian integration and Jacobian arguments therefore establish log-likelihood concavity. The participation-fee case follows by treating the purchased set as the discrete decision.

4. Flexibility through Mixtures of Kuhn-Tucker Models

We now examine whether mixtures of the KT models in Section 3.1 can approximate consumption distributions across varying prices, budgets, and assortments. We study this question in the single-budget setting without outside consumption.

4.1. A Limitation of Standard Mixed MDCEV

We first consider the standard MDCEV model, whose closed-form likelihood makes it attractive for empirical applications (Bhat 2008). Its utility is

$$v(q; \epsilon) = \sum_{j \in A} e^{\beta_j + \epsilon_j} \frac{\gamma_j}{\alpha_j} \left[\left(1 + \frac{q_j}{\gamma_j} \right)^{\alpha_j} - 1 \right],$$

where β_j captures baseline tastes, $\gamma_j > 0$ and $\alpha_j < 1$ govern diminishing marginal utility, and ϵ_j are independent type-I extreme-value shocks. The case $\alpha_j = 0$ uses the logarithmic limit.

This utility is additively separable across goods, and each good has diminishing marginal utility. These properties impose a fundamental limitation: holding prices, assortment, and realized preferences fixed, a larger budget reduces the budget multiplier and therefore weakly increases the quantity of every good. Mixing MDCEV types preserves this restriction: each type consumes weakly more of each good as the budget increases. In practice, a household with a larger food budget may replace some instant-noodle meals with more expensive, freshly prepared meals. Interactions between goods can represent this response even when utility is concave. To illustrate, consider the quadratic–Gaussian model of Section 3.2 with two goods and scale its systematic utility by $M > 0$:

$$p = (1, 3)^\top, \quad a = M(19, 55)^\top, \quad B = M \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}.$$

Dividing the utility objective by M leaves the optimal choice unchanged and replaces the standard Gaussian shock ϵ with ϵ/M . Thus, as $M \rightarrow \infty$, the shocks vanish, and the optimal quantities converge to

Budget	Good 1	Good 2
4	1	1
5	1/2	3/2

In this example, consumption of good 1 falls from 1 to $1/2$ as the budget rises from 4 to 5. A standard MDCEV mixture cannot approximate these choices at both budgets arbitrarily closely: every type’s consumption weakly increases with the budget. A positive approximation error therefore remains, regardless of how many types the mixture contains.

4.2. Approximating Demand across Contexts

We next show that mixtures of the KT models in Section 3.1 can approximate demand generated by continuous, increasing utilities, even when those utilities are nonconcave.

Consider a compact set $\mathcal{C}$ of contexts $c = (p, m, A)$, with strictly positive prices and budgets and finitely many nonempty assortments. Using the feasible set $K(c)$ from (1), let $\mathcal{X} = \bigcup_{c \in \mathcal{C}} K(c)$ denote the set of all feasible quantity bundles across these contexts. The bounds on prices and budgets ensure that $\mathcal{X}$ is compact. Let $v(q, \omega)$ denote the target utility whose induced consumption choices we seek to approximate, where ω captures preference heterogeneity and utility shocks drawn from a fixed population distribution. Formally, $v(\cdot, \omega)$ is a measurable random element of $C(\mathcal{X})$, the space of continuous functions on $\mathcal{X}$ equipped with the uniform norm. For almost every ω , assume that $v(q, \omega)$ is strictly increasing in quantities within each feasible set $K(c)$. It need not be concave. At each fixed context $c \in \mathcal{C}$, assume that the consumer’s optimal choice is unique with probability one, and write $q^*(c, \omega) = \arg \max_{q \in K(c)} v(q, \omega)$. Let $P^*(c)$ denote the joint distribution of these chosen quantities across consumers in context c .

To approximate these distributions, we use T consumer types. Type t has a smooth, strongly concave systematic utility v_t and chooses quantities according to the Gaussian-shock model in Section 3.1:

$$q_t(c, \epsilon) = \arg \max_{q \in K(c)} \{v_t(q) + \epsilon^\top q\}, \quad \epsilon \sim N(0, I_J).$$

Let $P_t(c)$ denote the resulting joint quantity distribution for type t . If w_t is its population share, with $w_t \geq 0$ and $\sum_{t=1}^T w_t = 1$, the mixture’s quantity distribution is $\hat{P}(c) = \sum_{t=1}^T w_t P_t(c)$. The same utilities v_t and shares w_t are used across all contexts in $\mathcal{C}$.

We measure approximation error using the Wasserstein distance. For two quantity distributions P and Q , define

$$W_1(P, Q) = \inf_{\substack{X \sim P \\ Y \sim Q}} \mathbb{E}[\|X - Y\|_2].$$

Here X and Y are random quantity vectors, and the infimum is over all joint distributions of (X, Y) with marginal distributions P and Q .

THEOREM 2 (Uniform approximation of choice distributions). *Under the preceding assumptions, for every $\varepsilon > 0$, there exists a finite mixture of the KT models in Section 3.1 such that*

$$\sup_{c \in \mathcal{C}} W_1(P^*(c), \hat{P}(c)) < \varepsilon.$$

Thus, one mixture can approximate the target quantity distribution to arbitrary accuracy across all prices, budgets, and assortments in $\mathcal{C}$. The mixture can be constructed using strongly concave polynomial utilities. Each type’s utility is a linear combination of fixed polynomial terms, with coefficients restricted to ensure strong concavity. These utilities therefore have the linear-in-parameters form in (3). Matzkin and Richter (1991) show that a consumer’s optimal decisions at finitely many prices and budgets can be reproduced by a strictly concave utility, even when the original utility is nonconcave. For each realization of the target utility, we use their result to construct a concave utility that reproduces these decisions. We then approximate this constructed concave utility with a strongly concave polynomial. By considering decisions under more contexts and refining the polynomial approximation, we can make the quantity error uniformly small across $\mathcal{C}$. Consumers with similar constructed utilities are represented by finitely many types, with their population shares determining the mixture weights. Finally, scaling the component utilities makes the effect of standard Gaussian shocks arbitrarily small. Appendix B provides the complete construction.

In practice, we specify a component utility family and a finite number of types, then use expectation-maximization (EM) to estimate their utility coefficients and population shares. By Theorem 1, the M-step remains a convex optimization problem. Under standard regularity conditions, the log likelihood converges monotonically to its value at a stationary point (Wu 1983).

5. Numerical Study

The numerical study evaluates the computational gains from our likelihood reformulation and the empirical fit of KT models and their mixtures.

5.1. Computational Performance

We compare computational performance for quadratic utility and a nonlinear specification that adds logarithmic and quartic terms:

$$v_{\theta}(q) = a^{\top} q - \frac{1}{2} q^{\top} B q + \sum_{j=1}^J \alpha_j \log(1 + q_j) - \frac{1}{4} \sum_{j=1}^J \kappa_j q_j^4.$$

The quadratic specification sets $\alpha = \kappa = 0$; the nonlinear specification also estimates α and κ . Both include linear outside utility $h_{\eta}(y) = \eta y$, with η estimated jointly with the inside-utility coefficients. We impose $B \succeq 10^{-5} I_J$ and $\alpha, \kappa \geq 0$ during generation and estimation, ensuring strong concavity.

We consider $J = 4, 10$, and 50 , with training samples of $400, 800$, and $4,000$ observations, respectively. The four- and ten-good designs use dense interaction matrices. For fifty goods, we impose $B_{jk} = 0$ whenever $|j - k| > 2$ during both generation and estimation. We generate ten independent instances for each utility specification and dimension, with prices, budgets, and assortments varying across observations. Appendix C.1.1 gives the generating distributions and coefficient bounds. We compare three implementations of maximum-likelihood estimation and an optimal-transport estimator:

1. **Dutt-based MLE.** We use the approximation of Dutt (1976) to evaluate multivariate Gaussian probabilities in the traditional KT likelihood of Wales and Woodland (1983).

2. **GHK-based MLE.** We use the Geweke–Hajivassiliou–Keane simulator to evaluate multivariate Gaussian probabilities in the traditional KT likelihood, following its use in MDC likelihood estimation by Kim et al. (2002).

3. **One-dimensional MLE.** We evaluate the same likelihood and its gradient using the one-dimensional integral described in Section 3.4.

4. **Optimal transport.** We adapt the conditional optimal-transport dual formulation of Carlier et al. (2016) to estimate the coefficients of our utility specifications. Appendix C.1.2 gives the objective and implementation details.

All four methods use the same constrained BFGS procedure (Nocedal and Wright 2006), with a 600-second limit per instance. Let θ collect all estimated coefficients, including the outside-utility coefficient, and write $B(\theta)$ for the corresponding interaction matrix. Let $F(\theta)$ denote the negative average log likelihood or the transport objective. At iteration k , BFGS uses the local quadratic model $g_k^\top d + \frac{1}{2}d^\top M_k d$ to determine the update direction, where $g_k = \nabla F(\theta_k)$ and $M_k \succ 0$ approximates the Hessian of F using changes in successive gradients. To keep the updated matrix B positive definite, we minimize this model subject to $B(\theta_k + d) \succeq 10^{-5}I_J$ and the coefficient bounds $\underline{\theta} \leq \theta_k + d \leq \bar{\theta}$. Introducing an auxiliary variable t for the quadratic term gives the SDP

$$\begin{aligned} \min_{d,t} \quad & g_k^\top d + t \\ \text{subject to} \quad & \begin{pmatrix} 2t & d^\top \\ d & M_k^{-1} \end{pmatrix} \succeq 0, \\ & B(\theta_k + d) \succeq 10^{-5}I_J, \\ & \underline{\theta} \leq \theta_k + d \leq \bar{\theta}. \end{aligned}$$

We use the resulting direction in a line search to determine the step size.

Table 1 Average estimation time (seconds)

Goods	Utility	Dutt MLE	GHK MLE	One-dimensional MLE	Optimal transport
4	Quadratic	23.28	7.81	0.81	3.07
	Nonlinear	9.82	6.94	1.25	NA
10	Quadratic	NA	NA	5.60	174.06
	Nonlinear	NA	NA	7.47	NA
50	Quadratic	NA	NA	147.13	NA
	Nonlinear	NA	NA	295.74	NA

Notes: Numerical entries are average estimation times over ten instances. NA indicates that at least one instance exceeds the 600-second time limit or 2 GiB memory limit. Appendix C.1.3 reports the stopping criteria and completion counts.

Our method completes all 60 instances in Table 1. For fifty goods, estimation takes 147.13 seconds on average under quadratic utility and 295.74 seconds under nonlinear utility. Dutt-based and GHK-based MLE

complete all ten instances only in the four-good configurations. Optimal transport completes the ten-good quadratic instances in 174.06 seconds on average, compared with 5.60 seconds for our method, but fails to complete all instances even in the four-good nonlinear configuration.

5.2. Empirical Performance

We compare the empirical fit of KT models and MDCEV on two datasets covering monetary and time allocation. Complete Journey (CJ) records household-week retail expenditure by product category ($J = 29$, $n = 66,981$).¹ We set the budget equal to observed total weekly expenditure, so every observation exhausts the budget. KuaiRand records user-day viewing time by video category ($J = 39$, $n = 27,830$) (Gao et al. 2022). We use a daily budget of 24 hours, and every observation leaves time unspent.

We estimate MDCEV, quadratic KT, and quadratic KT augmented with logarithmic utility terms, along with their four-type mixtures. The logarithmic specification uses three fixed scales per good. For KuaiRand, the KT models allow unspent time directly, while MDCEV includes an outside good representing the remainder of the day. All six models use the same 80:20 split of households or users into training and test sets within each dataset. Each fit uses a one-hour time limit. Table 2 reports average negative log likelihood (NLL) per observation on the training and test samples. Appendix C.2 provides the data construction, estimation details, and running times.

Table 2 Average negative log likelihood

Model	Complete Journey		KuaiRand	
	Training	Test	Training	Test
<i>Single type</i>				
MDCEV	31.6367	30.6232	51.1326	53.2266
Quadratic	31.8107	30.8256	75.5867	79.2835
Quadratic + log	31.2864	30.5933	49.3295	51.8714
<i>Four types</i>				
MDCEV	31.3568	30.3549	49.1618	51.5425
Quadratic	31.1532	30.2398	65.3585	68.8995
Quadratic + log	30.9916	30.1852	47.3723	50.0696

Notes: Lower values indicate better fit. Bold values mark the lowest NLL within each group of single-type or four-type models.

Among the single-type models, quadratic utility fits worse than MDCEV on both datasets, particularly in KuaiRand. Adding logarithmic terms reverses this ranking: Quadratic + log achieves lower training and test NLL than MDCEV in every application. Thus, a simple extension within our concave KT class improves empirical fit across these applications.

¹ <https://bradleyboehmke.github.io/completejourney/>

Allowing four types improves training and test fit for every utility specification on both datasets. The four-type Quadratic + log model achieves the lowest NLL in every column of Table 2, outperforming four-type MDCEV in each application. These results show that our framework can estimate heterogeneous KT models with better empirical fit than mixed MDCEV with both exhausted and unspent budgets.

6. Conclusion

By making general KT demand models estimable at scale, this paper helps connect economic decision analysis with consumer data. Researchers can calibrate the same utility models used to derive theoretical insights about pricing, assortment, and competition, and examine whether those insights persist under the estimated preferences. This gives researchers more freedom to choose utility specifications for the economic behavior they need to describe.

Our approximation results show that consumer heterogeneity alone does not ensure a flexible demand model. As mixed MDCEV demonstrates, increasing the number of types cannot eliminate behavioral restrictions shared by every component. The empirical findings reinforce the importance of enriching component utilities alongside introducing additional consumer types. KT mixtures accommodate such enrichment while retaining interpretable consumer types whose decisions can be analyzed through KKT conditions.

Future research could develop guidance for choosing utility complexity and the number of consumer types, and study how estimation errors affect firms' decisions. Although our analysis draws on the convexity theory of chance-constrained optimization, our estimation procedure has not exploited the specialized algorithms developed in that literature, such as scenario grouping and decomposition (Deng et al. 2021). Exploring how these algorithms can be used for maximum-likelihood estimation is a promising direction.

References

- Abdallah, T., A. Braverman, W. Gu. 2026. Multipurchase assortment optimization under a general random utility model. *Operations Research* **74**(5) 2772–2796. doi:10.1287/opre.2025.1702.
- Aouad, A., A. Désir. 2026. Representing random utility choice models with neural networks. *Management Science* **72**(8) 6686–6701. doi:10.1287/mnsc.2023.02189.
- Aouad, A., A. N. Elmachtoub, K. J. Ferreira, R. McNellis. 2023. Market segmentation trees. *Manufacturing & Service Operations Management* **25**(2) 648–667. doi:10.1287/msom.2023.1195.
- Bai, Y., J. Feldman, D. Segev, H. Topaloglu, L. Wagner. 2024. Assortment optimization under the multi-purchase multinomial logit choice model. *Operations Research* **72**(6) 2631–2664. doi:10.1287/opre.2023.2463.
- Bhat, C. R. 2008. The multiple discrete-continuous extreme value (MDCEV) model: Role of utility function parameters, identification considerations, and model extensions. *Transportation Research Part B* **42** 274–303.
- Bhat, C. R. 2022. A new closed-form two-stage budgeting-based multiple discrete-continuous model. *Transportation Research Part B* **164** 162–192. Author manuscript.

- Bienstock, D., M. Chertkov, S. Harnett. 2014. Chance-constrained optimal power flow: Risk-aware network control under uncertainty. *SIAM Review* **56**(3) 461–495. doi:10.1137/130910312.
- Blanchet, J., G. Gallego, V. Goyal. 2016. A Markov chain approximation to choice modeling. *Operations Research* **64**(4) 886–905. doi:10.1287/opre.2016.1505.
- Bonami, P., M. A. Lejeune. 2009. An exact solution approach for portfolio optimization problems under stochastic and integer constraints. *Operations Research* **57**(3) 650–670. doi:10.1287/opre.1080.0599.
- Carlier, G., V. Chernozhukov, A. Galichon. 2016. Vector quantile regression: An optimal transport approach. *The Annals of Statistics* **44**(3) 1165–1192. doi:10.1214/15-AOS1401.
- Caro, F., V. Martínez-de Albéniz. 2012. Product and price competition with satiation effects. *Management Science* **58**(7) 1357–1373. doi:10.1287/mnsc.1110.1489.
- Chen, Y.-C., V. V. Mišić. 2022. Decision forest: A nonparametric approach to modeling irrational choice. *Management Science* **68**(10) 7090–7111. doi:10.1287/mnsc.2021.4256.
- Şimşek, A. S., H. Topaloglu. 2018. Technical note—an expectation-maximization algorithm to estimate the parameters of the Markov chain choice model. *Operations Research* **66**(3) 748–760. doi:10.1287/opre.2017.1692.
- Deng, Y., H. Jia, S. Ahmed, J. Lee, S. Shen. 2021. Scenario grouping and decomposition algorithms for chance-constrained programs. *INFORMS Journal on Computing* **33**(2) 757–773. doi:10.1287/ijoc.2020.0970.
- Dubin, J. A., D. L. McFadden. 1984. An econometric analysis of residential electric appliance holdings and consumption. *Econometrica* **52**(2) 345–362. Article.
- Dutt, J. E. 1976. Numerical aspects of multivariate normal probabilities in econometric models. *Annals of Economic and Social Measurement*, vol. 5(4). NBER, 547–561. NBER chapter.
- El Housni, O., S. Sun, R. Udmani. 2025. A Markovian approach for cross-category complementarity in choice modeling. doi:10.48550/arXiv.2508.18566. ArXiv preprint arXiv:2508.18566.
- Farias, V. F., S. Jagabathula, D. Shah. 2013. A nonparametric approach to modeling choice with limited data. *Management Science* **59**(2) 305–322. doi:10.1287/mnsc.1120.1610.
- Federgruen, A., M. Hu. 2015. Multi-product price and assortment competition. *Operations Research* **63**(3) 572–584. doi:10.1287/opre.2015.1380.
- Feng, G., X. Li, Z. Wang. 2017. Technical note—on the relation between several discrete choice models. *Operations Research* **65**(6) 1516–1525. doi:10.1287/opre.2017.1602.
- Ferreira, K. J., J. Goh. 2021. Assortment rotation and the value of concealment. *Management Science* **67**(3) 1489–1507. doi:10.1287/mnsc.2020.3610.
- Gao, Chongming, Shijun Li, Yuan Zhang, Jiawei Chen, Biao Li, Wenqiang Lei, Peng Jiang, Xiangnan He. 2022. KuaiRand: An unbiased sequential recommendation dataset with randomly exposed videos. *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*. 3953–3957. doi:10.1145/3511808.3557624.

- Garikapati, V. M., R. Sidharthan, R. M. Pendyala, C. R. Bhat. 2014. Characterizing household vehicle fleet composition and count by type in integrated modeling framework. *Transportation Research Record* **2429**(1) 129–137. doi:10.3141/2429-14.
- Gurvich, I., J. Luedtke, T. Tezcan. 2010. Staffing call centers with uncertain demand forecasts: A chance-constrained optimization approach. *Management Science* **56**(7) 1093–1115. doi:10.1287/mnsc.1100.1173.
- Huh, W. T., H. Li. 2022. Technical note—optimal pricing under multiple-discrete customer choices and diminishing return of consumption. *Operations Research* **70**(2) 905–917. doi:10.1287/opre.2021.2146.
- Jagabathula, S., P. Rusmevichientong, A. Venkataraman, X. Zhao. 2024. Estimating large-scale tree logit models. *Operations Research* **72**(1) 257–276. doi:10.1287/opre.2023.2479.
- Jagabathula, S., L. Subramanian, A. Venkataraman. 2020. A conditional gradient approach for nonparametric estimation of mixing distributions. *Management Science* **66**(8) 3635–3656. doi:10.1287/mnsc.2019.3373.
- Jasin, S., C. Lyu, S. Najafi, H. Zhang. 2024. Assortment optimization with multi-item basket purchase under multivariate MNL model. *Manufacturing & Service Operations Management* **26**(1) 215–232. doi:10.1287/msom.2021.0526.
- Jiang, Y., C. Shi, S. Shen. 2019. Service level constrained inventory systems. *Production and Operations Management* **28**(9) 2365–2389. doi:10.1111/poms.13060.
- Ke, C., R. Wang. 2022. Cross-category retailing management: Substitution and complementarity. *Manufacturing & Service Operations Management* **24**(2) 1128–1145. doi:10.1287/msom.2021.0968.
- Kim, J., G. M. Allenby, P. E. Rossi. 2002. Modeling consumer demand for variety. *Marketing Science* **21**(3) 229–250.
- Lee, L.-F., M. M. Pitt. 1986. Microeconomic demand systems with binding non-negativity constraints: The dual approach. *Econometrica* **54**(5) 1237–1242.
- Lei, X., B. Wan, S. Wang. 2023. Content rotation in the presence of satiation effects. Working paper. SSRN.
- Lewbel, A., L. Nesheim. 2019. Sparse demand systems: Corners and complements. CeMMAP Working Paper CWP45/19, Centre for Microdata Methods and Practice, Institute for Fiscal Studies. doi:10.1920/wp.cem.2019.4519.
- Lin, H., X. Li, L. Wu. 2026. Intra-category multi-choice preferences learning and assortment recommendation in e-commerce. *Production and Operations Management* **35**(1) 304–330. doi:10.1177/10591478251350853.
- Liu, C., H. Sun, C.-P. Teo, X. Jin. 2025. Pricing analytics for ancillary products: Learning from conversion data to optimize revenue. Working paper. SSRN.
- Mas-Colell, A. 1978. On revealed preference analysis. *Review of Economic Studies* **45**(1) 121–131.
- Matzkin, R. L., M. K. Richter. 1991. Testing strictly concave rationality. *Journal of Economic Theory* **53** 287–303.
- McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. P. Zarembka, ed., *Frontiers in Econometrics*. Academic Press, New York, 105–142.

- McFadden, D., K. Train. 2000. Mixed MNL models for discrete response. *Journal of Applied Econometrics* **15** 447–470.
- Millimet, D. L., R. Tchernis. 2008. Estimating high-dimensional demand systems in the presence of many binding non-negativity constraints. *Journal of Econometrics* **147** 384–395. doi:10.1016/j.jeconom.2008.09.013.
- Mondal, A., C. R. Bhat. 2021. A new closed form multiple discrete-continuous extreme value (MDCEV) choice model with multiple linear constraints. *Transportation Research Part B* **147** 42–66. doi:10.1016/j.trb.2021.03.005.
- Nemirovski, A., A. Shapiro. 2006. Convex approximations of chance constrained programs. *SIAM Journal on Optimization* **17**(4) 969–996. doi:10.1137/050622328.
- Nocedal, J., S. J. Wright. 2006. *Numerical Optimization*. 2nd ed. Springer, New York. doi:10.1007/978-0-387-40065-5.
- Palma, D., S. Hess. 2022. Extending the multiple discrete continuous (MDC) modelling framework to consider complementarity, substitution, and an unobserved budget. *Transportation Research Part B* **161** 13–35. doi:10.1016/j.trb.2022.04.005.
- Phaneuf, D. J., C. L. Kling, J. A. Herriges. 2000. Estimation and welfare calculations in a generalized corner solution model with an application to recreation demand. *Review of Economics and Statistics* **82**(1) 83–92. doi:10.1162/003465300558650.
- Pinjari, A. R., C. R. Bhat, S. Saxena, A. Mondal. 2024. Multiple discrete-continuous choice models: A reflective analysis and a prospective view. S. Hess, A. Daly, eds., *Handbook of Choice Modelling*, 2nd ed., chap. 17. Edward Elgar, 452–488. doi:10.4337/9781800375635.00025.
- Rochet, J.-C., P. Choné. 1998. Ironing, sweeping, and multidimensional screening. *Econometrica* **66**(4) 783–826. doi:10.2307/2999574.
- Satomura, T., J. Kim, G. M. Allenby. 2011. Multiple constraint choice models with corner and interior solutions. *Marketing Science* **30**(3) 481–490. August 2010 author manuscript consulted.
- Shapiro, A., D. Dentcheva, A. Ruszczyński. 2009. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia. Book.
- Shen, S., J. C. Smith, S. Ahmed. 2010. Expectation and chance-constrained models and algorithms for insuring critical paths. *Management Science* **56**(10) 1794–1814. doi:10.1287/mnsc.1100.1208.
- Singh, N., X. Vives. 1984. Price and quantity competition in a differentiated duopoly. *RAND Journal of Economics* **15**(4) 546–554.
- Sun, H. 2026. Learning correlated preferences via inverse optimization: Applications to bundle pricing. Working paper. SSRN.
- Tanner, R., D. Bolduc. 2012. The multiple discrete-continuous extreme value model (MDCEV) with fixed costs. MPRA Paper 41452, Munich Personal RePEc Archive. Working paper.
- Train, K. E. 2009. *Discrete Choice Methods with Simulation*. 2nd ed. Cambridge University Press. Chapter 5: Probit.

- Tulabandhula, T., D. Sinha, S. R. Karra, P. Patidar. 2023. Multi-purchase behavior: Modeling, estimation, and optimization. *Manufacturing & Service Operations Management* **25**(6) 2298–2313. doi:10.1287/msom.2020.0238.
- van Ryzin, G., G. Vulcano. 2015. A market discovery algorithm to estimate a general class of nonparametric choice models. *Management Science* **61**(2) 281–300. doi:10.1287/mnsc.2014.2040.
- Wales, T. J., A. D. Woodland. 1983. Estimation of consumer demand systems with binding non-negativity constraints. *Journal of Econometrics* **21** 263–285.
- Wu, C. F. J. 1983. On the convergence properties of the EM algorithm. *The Annals of Statistics* **11**(1) 95–103. doi:10.1214/aos/1176346060.
- Zhang, H., H. Piri, W. T. Huh, H. Li. 2023. Assortment optimization under multiple-discrete customer choices. Working paper. SSRN.

Online Appendix: Kuhn–Tucker Demand Models: Scalable Estimation and Universal Approximation

Appendix A: Proofs for Section 3

A.1. Convex Estimation

Proof of Theorem 1. Fix an observation (q_i, c_i) and its purchased set S_i . For the general utility in Section 3.1, define

$$t_{i,\theta} = -\nabla v_\theta(q_i), \quad H_\theta(q_i) = -\nabla^2 v_\theta(q_i).$$

Both are affine in θ because utility is linear in its coefficients. The proof follows the intuition in Section 3.3: the purchased-good conditions determine a density factor, while the unpurchased-good conditions give a probability of linear shock inequalities. We show that both factors are log-concave. Strong concavity ensures that the Jacobian within the density factor has the required log-determinant structure.

Likelihood reformulation. The purchased-good KKT conditions are $\epsilon_{ij} = t_{i,\theta,j} + \lambda_i p_{ij}$, and the unpurchased-good conditions are $\epsilon_{ij} \leq t_{i,\theta,j} + \lambda_i p_{ij}$. When the budget is exhausted, choose $r_i \in S_i$ and use the budget equality to determine q_{ir_i} from the other purchased quantities. The purchased-good equalities transform the shocks into these $|S_i| - 1$ free quantities and λ_i . Strong concavity makes this transformation one-to-one, and its Jacobian $C_i^\lambda(\theta)$, calculated below, does not depend on λ_i . Integrating out the unpurchased shocks and the multiplier gives

$$f_\theta(q_i | c_i) = C_i^\lambda(\theta) \int_0^\infty \prod_{j \in S_i} \phi(t_{i,\theta,j} + \lambda p_{ij}) \prod_{j \in A_i \setminus S_i} \Phi(t_{i,\theta,j} + \lambda p_{ij}) d\lambda.$$

Define

$$\bar{\lambda}_i(\theta) = -\frac{\sum_{j \in S_i} p_{ij} t_{i,\theta,j}}{\sum_{j \in S_i} p_{ij}^2}, \quad \sigma_i^2 = \frac{1}{\sum_{j \in S_i} p_{ij}^2}.$$

Completing the square in the purchased-good densities and substituting $\lambda = \bar{\lambda}_i(\theta) + \sigma_i z$ yields $f_\theta(q_i | c_i) = D_i(\theta) P_i(\theta)$, where

$$D_i(\theta) = C_i^\lambda(\theta) \sigma_i (2\pi)^{-(|S_i|-1)/2} \exp \left[-\frac{1}{2} \sum_{j \in S_i} (t_{i,\theta,j} + p_{ij} \bar{\lambda}_i(\theta))^2 \right]$$

and, with $Z_i \sim N(0, 1)$ independent of the unpurchased shocks,

$$P_i(\theta) = \Pr \left\{ \begin{array}{l} -\sigma_i Z_i \leq \bar{\lambda}_i(\theta), \\ \epsilon_{ij} - p_{ij} \sigma_i Z_i \leq t_{i,\theta,j} + p_{ij} \bar{\lambda}_i(\theta), \quad j \in A_i \setminus S_i \end{array} \right\}.$$

When the budget is unexhausted, $\lambda_i = 0$, and the density in the purchased quantities has the same decomposition with

$$D_i(\theta) = C_i(\theta) \prod_{j \in S_i} \phi(t_{i,\theta,j}), \quad P_i(\theta) = \prod_{j \in A_i \setminus S_i} \Phi(t_{i,\theta,j}).$$

As in (12), introducing $P_i(\theta) \geq e^{s_i}$ and maximizing $\sum_i [\log D_i(\theta) + s_i]$ gives an equivalent chance-constrained problem: at the optimum, $s_i = \log P_i(\theta)$.

The chance constraints. Each threshold defining $P_i(\theta)$ is affine in θ , including $\bar{\lambda}_i(\theta)$. The random variables have a coefficient-independent Gaussian distribution. The probability of these affine inequalities is therefore log-concave (Shapiro, Dentcheva, and Ruszczyński 2009, Theorem 4.39). This applies to both budget cases. Consequently, $s_i - \log P_i(\theta)$ is convex, so the chance constraints and the convex parameter domain Θ define a convex feasible region.

The density factor and its Jacobian. For an exhausted budget, order r_i last among purchased goods and define

$$V_i = \frac{\partial q_{iS_i}}{\partial q_{i,S_i \setminus \{r_i\}}} = \begin{pmatrix} I_{|S_i|-1} \\ -p_{i,S_i \setminus \{r_i\}}^T / p_{ir_i} \end{pmatrix}, \quad H_i(\theta) = V_i^\top H_{\theta,S_i S_i}(q_i) V_i.$$

Here $H_i(\theta)$ is the negative Hessian after substituting the budget equality. It is positive definite by strong concavity and affine in θ . The purchased-good transformation has derivative $[H_{\theta,S_i S_i}(q_i) V_i, p_{iS_i}]$. If e_{r_i} selects the last purchased coordinate, then

$$\begin{pmatrix} V_i^\top \\ e_{r_i}^\top \end{pmatrix} [H_{\theta,S_i S_i}(q_i) V_i, p_{iS_i}] = \begin{pmatrix} H_i(\theta) & 0 \\ e_{r_i}^\top H_{\theta,S_i S_i}(q_i) V_i & p_{ir_i} \end{pmatrix}.$$

The premultiplying matrix has determinant one. For an unexhausted budget, the purchased shocks are directly determined by the purchased quantities, with derivative $H_{\theta,S_i S_i}(q_i)$. Thus

$$C_i^\lambda(\theta) = p_{ir_i} \det H_i(\theta), \quad C_i(\theta) = \det H_{\theta,S_i S_i}(q_i). \quad (\text{A.1})$$

Both Jacobians are log-concave: their matrices are positive definite and affine in θ , and the price factor is independent of θ . Empty determinants equal one, covering observations with no freely varying quantities. The normalized-difference Jacobian in (4) is likewise $C_i^\Delta(\theta) = C_i^\lambda(\theta) / \prod_{j \in S_i} p_{ij}$, since normalizing the purchased shocks divides the determinant by this product.

For an exhausted budget,

$$\log D_i(\theta) = \log C_i^\lambda(\theta) - \frac{1}{2} \sum_{j \in S_i} (t_{i,\theta,j} + p_{ij} \bar{\lambda}_i(\theta))^2 + \log \sigma_i - \frac{|S_i|-1}{2} \log(2\pi).$$

The squared terms have affine arguments, and the last two terms are constant in θ . Hence $\log D_i(\theta)$ is concave. The same conclusion holds for an unexhausted budget, where the purchased-good log densities are negative squares of $t_{i,\theta,j}$. The chance-constrained objective is therefore concave. Equivalently, $\log f_\theta(q_i | c_i) = \log D_i(\theta) + \log P_i(\theta)$ is concave for each observation. Summing over observations proves Theorem 1.

Log-concave shocks. For the extension stated after the theorem, let f_{A_i} be a log-concave joint shock density. Retain the unpurchased shocks $u = (u_j)_{j \in A_i \setminus S_i}$ as integration variables. For an exhausted budget, the same change of variables gives

$$f_\theta(q_i | c_i) = C_i^\lambda(\theta) \int_{\lambda \geq 0} \int_{u_j \leq t_{i,\theta,j} + \lambda p_{ij}, j \in A_i \setminus S_i} f_{A_i}((t_{i,\theta,j} + \lambda p_{ij})_{j \in S_i}, u) du d\lambda,$$

where the density's coordinates are ordered with purchased goods first. The integrand, including the indicator of the integration region, is jointly log-concave in (θ, λ, u) : its density argument is affine and its restrictions are linear. Integrating out (λ, u) preserves log-concavity (Shapiro, Dentcheva, and Ruszczyński 2009, Theorems 4.15 and 4.26). The Jacobian is log-concave by (A.1). For an unexhausted budget, set $\lambda = 0$, omit its integral, and use $C_i(\theta)$. This proves the stated extension. $\square$

A.2. Population Identification

Proof of Proposition 1. Fix a context c and work in its $|A|$ available coordinates. Suppose two admissible utilities v_1 and v_2 generate the same quantity distribution under the known standard Gaussian shock law. We first show that they choose the same quantities for each shock. We then use interior choices to recover their marginal utilities.

From equal distributions to equal choices. Let e be a possible shock realization. For model $k \in \{1, 2\}$, denote its maximized utility by $W_k(e)$ and its chosen quantities by $q_k(e)$:

$$W_k(e) = \max_{q \in K(e)} \{v_k(q) + e^\top q\}, \quad q_k(e) = \arg \max_{q \in K(e)} \{v_k(q) + e^\top q\}.$$

For any fixed allocation, utility is affine in e . Its maximum over allocations, W_k , is therefore convex. Strong concavity of v_k makes $q_k(e)$ unique, and differentiating the maximized value gives $\nabla W_k(e) = q_k(e)$. Thus each model maps shocks to quantities through the gradient of a convex function.

We now apply the uniqueness result in Carlier, Chernozhukov, and Galichon (2016, Theorem 2.1). For a fixed absolutely continuous source distribution and a fixed target distribution, a gradient of a convex function that maps the source to the target is unique almost everywhere under the source law. In our setting, the source is the known Gaussian shock distribution and the target is the common quantity distribution. Hence $q_1(e) = q_2(e)$ for almost every shock. The theorem permits discrete or mixed target distributions, so zero purchases and binding-budget observations are included.

To use first-order conditions at any interior allocation, we need this equality for every shock value. Strong concavity makes the choice maps continuous. In fact, their optimality inequalities give

$$\mu \|q_k(e) - q_k(e')\|^2 \leq (e - e')^\top (q_k(e) - q_k(e')),$$

so $\|q_k(e) - q_k(e')\| \leq \|e - e'\|/\mu$. If the two maps differed at some shock, continuity would make them differ on a neighborhood of it. Every such neighborhood has positive Gaussian probability, contradicting their almost-everywhere equality. Therefore $q_1(e) = q_2(e)$ for every e .

From equal choices to equal marginal utilities. Choose any allocation q with every available quantity positive and $p^\top q < m$. Set $e = -\nabla v_1(q)$. The first model's marginal utilities are then zero at q , and strong concavity makes q its unique optimum. The second model makes the same choice for this shock. Since the allocation is interior, its first-order conditions give $\nabla v_2(q) + e = 0$, and hence $\nabla v_2(q) = \nabla v_1(q)$. This holds at every interior allocation. The interior is convex, so integrating the zero gradient of $v_2 - v_1$ along any segment shows that their difference is constant there. Continuity extends the same constant to all of $K(e)$. Conversely, a constant utility difference on $K(e)$ leaves the maximizer unchanged for every shock, and hence leaves the quantity distribution unchanged. This proves Proposition 1. $\square$

A.3. Outside Consumption

Proof of Corollary 1. For the outside utility in Section 3.5.1, the features are twice continuously differentiable for positive consumption. At zero, use their right-limit values, allowing utility to be minus infinity. At an observation with $y = m - p^\top q > 0$, the inverse gradient and negative Hessian of the reduced utility in (13) are

$$t_{\theta, \eta}(q) = h'_\eta(m - p^\top q)p - \nabla v_\theta(q), \quad H_{\theta, \eta}(q) = -\nabla^2 v_\theta(q) - h''_\eta(m - p^\top q)pp^\top. \quad (\text{A.2})$$

Both are affine in (θ, η) . Concavity of h_η makes the additional negative-Hessian term positive semidefinite, preserving the inside utility's positive curvature bound. Applying Appendix A.1 gives the likelihood in (11) with these derivatives and proves its joint log-concavity.

For an observation with $y = 0$, define

$$g_\eta = h'_\eta(0+) = \sup_{0 < y \leq m} h'_\eta(y) \in [0, \infty]. \quad (\text{A.3})$$

Concavity gives the equality. Each $h'_\eta(y)$ is affine in η , so g_η is a convex extended-real function. If $g_\eta < \infty$, optimality requires the total-budget multiplier to satisfy $\lambda \geq g_\eta$; transferring a unit of resource from inside consumption to the outside option cannot improve utility. Purchased shocks still satisfy $\epsilon_j = -[\nabla v_\theta(q)]_j + \lambda p_j$, with the corresponding upper bounds for unpurchased shocks. The likelihood is consequently (6), with $(Bq - a)_j$ replaced by $-\nabla v_\theta(q)_j$ and the lower integration limit replaced by g_η .

Outside consumption stays at zero along this face, so outside utility contributes no curvature in its freely varying quantity coordinates. The Jacobian therefore uses the inside utility's negative Hessian, and its log determinant is concave by (A.1). The integration region $\{(\eta, \lambda) : \lambda \geq g_\eta\}$ is convex, and the Gaussian integrand is jointly log-concave in (θ, λ) . Integration preserves log-concavity, proving the result for these boundary observations. When outside utility has finite derivatives at zero, shifting the multiplier by g_η recovers (6) for the reduced utility.

If $g_\eta = \infty$, the integration region is empty and the boundary likelihood is zero. Reducing a positive inside quantity slightly then creates outside consumption whose marginal gain exceeds the finite inside loss. An outside utility with $h_\eta(0) = -\infty$ also excludes this boundary. For a fixed log-concave shock density, retain the unpurchased shocks as integration variables; the same affine shock expressions and convex multiplier region give the same result. Combining all quantity faces proves the corollary. $\square$

A.4. Multiple Resource Constraints

Proof of Corollary 2. We first derive the multiplier formula for independent binding resource constraints, following the change of variables in Section 3.5.2. We then establish convexity when binding constraints are dependent.

Independent binding constraints. Fix an observation and suppress its index. Let R index the binding resources, S the purchased goods, and $I = A \setminus S$. Write $C = D_{R,S}$, $t_\theta(q) = -\nabla v_\theta(q)$, and $H_\theta = -\nabla^2 v_\theta(q)$. Suppose the rows of C are independent. Choose $L \subseteq S$ with $|L| = |R|$ such that $C_L = D_{R,L}$ is invertible, and put $Q = S \setminus L$. The binding equalities determine $q_L = C_L^{-1}(b_R - C_Q q_Q)$. Ordering Q first, the derivative of q_S with respect to the free quantities is

$$V = \begin{pmatrix} I_{|Q|} \\ -C_L^{-1} C_Q \end{pmatrix}.$$

The purchased shocks satisfy $\epsilon_S = t_{\theta,S} + C^\top \lambda$. Their derivative with respect to (q_Q, λ) is $[H_{\theta,SS} V, C^\top]$. Premultiplying by the determinant-one matrix with rows $V^\top$ and the selectors of L gives a block triangular matrix with diagonal blocks $V^\top H_{\theta,SS} V$ and $C_L^\top$. Thus the Jacobian is $|\det C_L| \det(V^\top H_{\theta,SS} V)$. After integrating the unpurchased shocks, the likelihood is

$$f_\theta(q \mid D, b) = |\det C_L| \det(V^\top H_{\theta,SS} V) \int_{\mathbb{R}_+^{|R|}} \prod_{j \in S} \phi(t_{\theta,j} + (D_R^\top \lambda)_j) \prod_{j \in I} \Phi(t_{\theta,j} + (D_R^\top \lambda)_j) d\lambda. \quad (\text{A.4})$$

The log-Jacobian is concave in θ , and the integrand is jointly log-concave in (θ, λ) . Marginalization therefore proves log-likelihood concavity. There are $|R|$ integration variables. Differentiating the same integral gives the gradient without introducing additional integration variables; Gaussian tails provide local domination because C has full row rank. When R is empty, the multiplier integral is absent and the formula reduces to the unexhausted-budget likelihood.

Dependent binding constraints. Redundant multipliers would count the same shocks more than once. We instead parameterize the marginal-utility vectors that support the observed choice. Let $K_D = \{q \geq 0 : Dq \leq b\}$, working in available coordinates. At q , these vectors form the convex set

$$N_{K_D}(q) = \{n : n^\top (w - q) \leq 0 \text{ for every } w \in K_D\},$$

called the normal cone. The KKT conditions are equivalent to $\epsilon = t_\theta(q) + n$ for some $n \in N_{K_D}(q)$.

Let F be the part of the feasible set with the same binding constraints as q . Choose a set Q of freely varying quantity coordinates on F ; the remaining quantities are affine functions of them. Choose a point $q_F \in F$ and write $q = q_F + V_F u$, where $u \in \mathbb{R}^{|Q|}$ records changes in those free quantities. Let N_F have orthonormal columns perpendicular to the columns of V_F , and set $C_F = \{z : N_F z \in N_{K_D}(q)\}$. The cone is constant on the relative interior of F , so C_F depends only on the observed binding constraints. A supporting shock is

$$\epsilon = -\nabla v_\theta(q_F + V_F u) + N_F z, \quad z \in C_F.$$

Its derivative with respect to (u, z) is $[H_\theta V_F, N_F]$. Premultiplication by $[V_F, N_F]^\top$ gives diagonal blocks $V_F^\top H_\theta V_F$ and $I_{|A|-|Q|}$, with a zero upper-right block. Since $|\det[V_F, N_F]| = \sqrt{\det(V_F^\top V_F)}$, the Jacobian is

$$J_F(\theta) = \frac{\det(V_F^\top H_\theta V_F)}{\sqrt{\det(V_F^\top V_F)}}. \quad (\text{A.5})$$

Strong concavity makes this change of variables one-to-one. The likelihood in the free quantity coordinates is therefore

$$f_\theta(q | D, b) = J_F(\theta) \int_{C_F} f_A(t_\theta(q) + N_F z) dz. \quad (\text{A.6})$$

The log-Jacobian is concave because its numerator is the determinant of a positive definite matrix affine in θ , while its denominator is independent of θ . The integral is log-concave by marginalization over the convex cone C_F . This proves convexity without requiring independent binding constraints. Empty determinants equal one; when no quantity varies freely, the formula gives the probability of the observed choice. The same argument applies to lower-dimensional feasible sets, as needed for the discrete extensions below. $\square$

A.5. Participation Fees and Binary Choices

Proof of Corollary 3. An observed choice must be optimal both within its discrete decision and against other discrete decisions. The first requirement gives the quantity–shock transformation from Appendix A.4; the second adds utility comparisons to the integration region.

Binary choices. Retain the $|A|$ available continuous goods and the K binary options in Section 3.5.3. Let ρ be their resource requirements and ξ their shocks. Conditional on $x \in \{0, 1\}^K$, the continuous feasible set is $K_x = \{q \geq 0 : p^\top q \leq m - \rho^\top x\}$.

Ignore binary vectors for which this set is empty. At the observed x_i , utility is smooth and strongly concave in q . Let F be the face of K_{x_i} containing q_i in its relative interior, and choose a set Q of free quantity coordinates on F . As in Appendix A.4, use the quantity-coordinate matrix V_F , its orthonormal complementary matrix N_F , and the cone $C_F = \{\zeta : N_F \zeta \in N_{K_{x_i}}(q_i)\}$.

Set $t_\theta = -\nabla_q v_\theta(q_i, x_i)$ and $H_\theta = -\nabla_q^2 v_\theta(q_i, x_i)$. Optimality conditional on the observed binary decision is equivalent to

$$\epsilon = t_\theta + N_F \zeta, \quad \zeta \in C_F. \quad (\text{A.7})$$

Changing from shocks to the $|Q|$ free quantities and the $|A| - |Q|$ cone coordinates has Jacobian $J_F(\theta)$ in (A.5). The binary shocks ξ remain integration variables.

To make the observed choice globally optimal, impose the utility comparison from Section 3.5.3 for every feasible (q, x) . After (A.7), these comparisons are

$$(t_\theta + N_F \zeta)^\top (q - q_i) + \xi^\top (x - x_i) \leq v_\theta(q_i, x_i) - v_\theta(q, x). \quad (\text{A.8})$$

For each fixed alternative, both sides are affine in (θ, ζ, ξ) . The alternative set is fixed by observed resource requirements and budget. Let $\mathcal{R}_i(\theta)$ collect the pairs (ζ, ξ) satisfying $\zeta \in C_F$ and every comparison in (A.8). Then

$$\{(\theta, \zeta, \xi) : \theta \in \Theta, (\zeta, \xi) \in \mathcal{R}_i(\theta)\}$$

is convex, being the intersection of the convex parameter domain, a fixed cone and affine half-spaces.

The global choice is unique almost surely. Maximum utility over the compact set of feasible (q, x) is a finite convex Lipschitz function of (ϵ, ξ) , hence differentiable almost everywhere. At a differentiability point, every maximizing (q, x) equals its gradient. Gaussian shocks have a density, so ties occur with probability zero. The change of variables therefore counts each choice once.

Let $\phi_{|A|}$ and ϕ_K be the standard Gaussian densities in their respective dimensions. With counting measure for the observed binary decision and the free coordinates for the continuous quantities, the joint choice density is

$$f_\theta(q_i, x_i | c_i) = J_F(\theta) \int_{\mathcal{R}_i(\theta)} \phi_{|A|}(t_\theta + N_F \zeta) \phi_K(\xi) d\zeta d\xi. \quad (\text{A.9})$$

The Jacobian is independent of the integrated variables. It is positive and log-concave in θ . The Gaussian integrand, including the indicator of the joint convex region, is jointly log-concave in (θ, ζ, ξ) . Marginalization therefore makes the integral log-concave in θ . Their product is log-concave, proving the binary-choice result.

Participation fees. Let F_j be the known fee for consuming good j . For each subset $S \subseteq A$, define the budget remaining after these fees and the corresponding quantity set by

$$m_S = m - \sum_{j \in S} F_j, \quad K^S = \{q \geq 0 : q_j = 0 \ (j \notin S), p^\top q \leq m_S\},$$

omitting empty sets. The fee-feasible set is exactly $\bigcup_{S \subseteq A} K^S$. A bundle in K^S pays no more in actual participation fees than the fees deducted in m_S , while every fee-feasible bundle belongs to the set indexed by its actual positive coordinates. This finite union is compact.

For an observed bundle, use its actual purchased set S_i and the face of K^{S_i} on which its positive quantities lie. Use the matrices V_F, N_F and cone coordinates ζ for this face as in Appendix A.4, with $t_\theta = -\nabla v_\theta(q_i)$. The same transformation and Jacobian apply. Because unpurchased quantities are fixed at zero within K^{S_i} , their shocks are unrestricted by optimality within this set. In particular, the baseline restrictions $\epsilon_j \leq -[\nabla v_\theta(q_i)]_j$ need not hold for goods with a positive entry fee.

Comparisons with other purchased sets supply the remaining restrictions:

$$(t_\theta + N_F \zeta)^\top (q - q_i) \leq v_\theta(q_i) - v_\theta(q) \quad \text{for every } q \in \bigcup_{S \subseteq A} K^S. \quad (\text{A.10})$$

These are again affine restrictions in (θ, ζ) . Applying (A.9) without the binary-shock variables proves log-concavity. Partitioning observations by their actual positive coordinates avoids counting a bundle more than once when it belongs to several of the sets K^S . Almost-sure global uniqueness follows from the same differentiability argument for maximum utility as a function of ϵ . Summing the observation log likelihoods on the convex coefficient domain completes the proof of the corollary. $\square$

Appendix B: Proofs for Section 4

Proof of Theorem 2. We use the target utilities and quantity distributions defined in Section 4.2. The proof first approximates each consumer’s choices by a strongly concave utility. It then replaces the population by finitely many representative polynomial utilities and adds standard Gaussian shocks.

B.1. Extending the target for the construction

The comparisons in Appendix B.3 use slightly different prices and budgets, which may lie outside $\mathcal{C}$. We therefore first extend the target utility beyond $\mathcal{X}$ and make its choices unique at each fixed comparison context. Both properties will follow from the construction below.

The set $\mathcal{X}$ is compact and downward closed: if $q \in \mathcal{X}$ and $0 \leq x \leq q$, then $x \in \mathcal{X}$. Positive prices bounded away from zero and bounded budgets give bounded quantities. Closedness follows by taking a convergent sequence of contexts and a subsequence with the same assortment.

For $q \in \mathbb{R}_+^J$, define

$$(Ev)(q; \omega) = \max\{v(x; \omega) : x \in \mathcal{X}, 0 \leq x \leq q\}. \quad (\text{B.1})$$

This maximum exists because its feasible set is nonempty and compact. Strict monotonicity within the original feasible sets implies $(Ev)(q; \omega) = v(q; \omega)$ for $q \in \mathcal{X}$. Increasing q enlarges the set in (B.1), so Ev is nondecreasing everywhere.

The extension is continuous. If $q_n \rightarrow q$, limits of feasible maximizers in (B.1) remain feasible at q , giving the upper bound on the limiting value. For the reverse bound, any $x \leq q$ in $\mathcal{X}$ is approximated by $x_n = \min\{x, q_n\}$, coordinate by coordinate. Downward closure makes x_n feasible, and continuity of v gives the desired lower bound. For any two continuous functions v, w on $\mathcal{X}$ and any compact box $\mathcal{B}$ containing $\mathcal{X}$, $\|Ev - Ew\|_{\infty, \mathcal{B}} \leq \|v - w\|_{\infty, \mathcal{X}}$. Thus the extension also preserves measurability as a continuous random function.

We need unique choices at the nearby budgets used later. Let η be independent of the target utility and uniform on $[1, 2]^J$, and define

$$v_\alpha(q; \omega, \eta) = (Ev)(q; \omega) + \alpha \eta^\top q, \quad \alpha > 0. \quad (\text{B.2})$$

This auxiliary utility is continuous and strictly increasing. At any fixed context and target realization, the value function $a \mapsto \max_{q \in K(c)} \{(Ev)(q; \omega) + a^\top q\}$ is finite, convex and Lipschitz. Every maximizing quantity is a subgradient: evaluating that quantity at another coefficient vector gives the subgradient inequality. At points where the value function is differentiable, the maximizer is therefore unique. A finite convex function is differentiable almost everywhere, and $\alpha \eta$ has a density. Hence the auxiliary choice is almost surely unique at each fixed positive-price, positive-budget context. We next show that the auxiliary choices approximate the target choices. Let $q_\alpha(c)$ denote the maximizer under v_α , suppressing its dependence on (ω, η) . Compare it with $q^*(c, \omega)$ for the same consumer; expectations below include both ω and η . Put $R = \max_{q \in \mathcal{X}} \sum_j q_j$. On $\mathcal{X}$, $0 \leq v_\alpha(q) - v(q) \leq 2\alpha R$. We claim that

$$\sup_{c \in \mathcal{C}} \mathbb{E} \|q_\alpha(c) - q^*(c, \omega)\|_2 \rightarrow 0 \quad \text{as } \alpha \downarrow 0. \quad (\text{B.3})$$

If (B.3) failed, there would be sequences $\alpha_n \downarrow 0$ and $c_n \in \mathcal{C}$ whose expected errors stay above a positive constant. Compactness gives a subsequence with $c_n \rightarrow c_\infty \in \mathcal{C}$ and a constant assortment. Choose this deterministic subsequence before considering the random utilities. The feasible sets converge as well: any bundle feasible at c_∞ can be scaled down slightly to become feasible at c_n , while every limit of feasible bundles remains feasible at c_∞ . On the

probability-one event of target uniqueness at c_∞ , continuity and the uniform utility-error bound imply that both maximizing bundles at c_n converge to that same target choice. Distances are bounded by the diameter of $\mathcal{X}$, so dominated convergence contradicts the positive expected error. This proves (B.3). Fix α so that the error in (B.3) is less than $\varepsilon/5$. We now approximate the choices of this auxiliary population.

B.2. Reproducing finitely many choices with a strongly concave utility

For each assortment, enlarge the context set slightly to an open set of positive prices and budgets, keeping prices bounded away from zero and budgets bounded above. All feasible quantities in these enlarged sets then fit inside a compact box $\mathcal{B}$. Choose a countable sequence $D = \{c_1, c_2, \dots\}$ that is dense in the enlarged context sets. Each context in D has a unique auxiliary choice with probability one; their countable intersection also has probability one. Fix an auxiliary consumer in this intersection and write $x_r = q_\alpha(c_r)$. We will construct a strongly concave utility that reproduces its first n choices.

At the first n contexts, the chosen bundles exhaust their budgets. Their revealed-preference rankings have no cycles because each choice uniquely maximizes the same utility. To use the ordinary-budget formulation of Matzkin and Richter (1991, Theorem 2), replace the prices of unavailable goods at each observation by sufficiently high finite prices. For this finite collection of bundles, these prices make every bundle using an unavailable good unaffordable and preserve all other affordability comparisons. Denote the resulting price vectors by $\tilde{p}_r$. On the original feasible set, $\tilde{p}_r^\top q = p_r^\top q$.

The strict Afriat inequalities in that theorem provide utility levels u_r and multipliers $\lambda_r > 0$ such that $u_s < u_r + \lambda_r \tilde{p}_r^\top (x_s - x_r)$ when $x_s \neq x_r$.

Repeated bundles share a utility level, while their budget rows and multipliers are retained. There are only finitely many strict inequalities. We can therefore choose $\delta_n > 0$ small enough that

$$u_s \leq u_r + \lambda_r \tilde{p}_r^\top (x_s - x_r) - \delta_n \|x_s - x_r\|_2^2 \quad \text{for all } r, s \leq n. \quad (\text{B.4})$$

Choose it smaller, if necessary, so that each function in the following minimum is increasing on $\mathcal{B}$:

$$V_n(q) = \min_{r \leq n} \{u_r + \lambda_r \tilde{p}_r^\top (q - x_r) - \delta_n \|q - x_r\|_2^2\}. \quad (\text{B.5})$$

Adding $\delta_n \|q\|_2^2$ to $V_n(q)$ leaves a minimum of affine functions, which is concave. Thus V_n is $2\delta_n$ -strongly concave. Equation (B.4) gives $V_n(x_r) = u_r$. For any different $q \in K(c_r)$, its r -th supporting expression gives $V_n(q) \leq u_r - \delta_n \|q - x_r\|_2^2 < u_r$.

Therefore V_n uniquely reproduces all n recorded choices.

Choose these utilities measurably across consumers as follows. Fix enumerations of the positive integers and rational parameter tuples. Choose the first sufficiently large prices for unavailable goods, then the first tuple $(u_r, \lambda_r, \delta_n)$ satisfying the strict inequalities for distinct bundles and strict monotonicity on $\mathcal{B}$. Such tuples exist because these finitely many restrictions have strict slack. Repeated bundles share one utility-level variable. The choices x_r are measurable, so this rule gives a measurable random utility V_n .

B.3. Controlling choices at all contexts

The utility V_n agrees with the auxiliary consumer at the first n recorded contexts. To obtain the theorem, we must also control choices at contexts not yet recorded. Let $q_{V_n}(c)$ maximize V_n over $K(c)$. We show

$$\sup_{c \in \mathcal{C}} \mathbb{E} \|q_{V_n}(c) - q_\alpha(c)\|_2 \longrightarrow 0. \quad (\text{B.6})$$

This uses increasingly dense budget comparisons, following the strategy of Mas-Colell (1978); the argument below supplies the uniform expected-error conclusion needed here.

If (B.6) failed, there would be contexts c_n with expected errors bounded away from zero. Choose a deterministic subsequence with $c_n \rightarrow c^* = (p^*, m^*, A^*)$ and the same assortment. On the probability-one event of auxiliary uniqueness at this fixed c^* and every context in D , consider one consumer. Write $x^* = q_\alpha(c^*)$. Continuity of utility and of the feasible sets implies $q_\alpha(c_n) \rightarrow x^*$.

If the rationalizer choices fail to converge to x^* , take a further subsequence with $q_{V_n}(c_n) \rightarrow x \neq x^*$. Both x and x^* exhaust the limiting budget because the auxiliary utility and every rationalizer are increasing. On the available coordinates, set $d = x - x^*$, and for small $s > 0$ define

$$p_s = p^* - sd, \quad m_s = p_s^\top x^* - \frac{s}{2} \|d\|_2^2. \quad (\text{B.7})$$

For sufficiently small s , this context remains in the enlarged set chosen in Appendix B.2. The competing bundle x is strictly affordable there, since $p_s^\top x = m_s - \frac{s}{2} \|d\|_2^2$.

Every auxiliary maximizer at sufficiently close contexts lies near x^* , by uniqueness at c^* and continuity. A bundle y feasible at the tilted budget satisfies $(p^*)^\top y \leq m^* + s d^\top (y - x^*) - \frac{s}{2} \|d\|_2^2$. If $\|y - x^*\|_2 < \|d\|_2/4$, the right side is strictly below m^* . Thus every auxiliary optimum at this tilted budget is strictly affordable at the original limiting budget.

Choose a context $c_r \in D$ sufficiently close to the tilted context to preserve both strict affordability inequalities and the separation from x . Its auxiliary choice y_r is unique. For large n , the bundle $q_{V_n}(c_n)$ is affordable at c_r , while y_r is affordable at c_n . Since $n \geq r$, the same strictly concave rationalizer chooses y_r at c_r and a distinct bundle at c_n . Unique optimality would require each bundle to have greater utility than the other, a contradiction.

The rationalizer choices therefore converge to x^* on the selected probability-one event. Bounded quantities then give convergence of the expected errors to zero, contradicting the selected sequence. This proves (B.6).

B.4. Obtaining finitely many types

Fix n so that the error in (B.6) is less than $\varepsilon/5$. We now group consumers whose constructed utilities are uniformly close on $\mathcal{B}$.

First choose $\gamma > 0$ small enough that consumers with $2\delta_n < \gamma$ have arbitrarily small probability. This is possible because $\delta_n > 0$ for every constructed utility. The law of V_n is a Borel probability measure on the separable complete space $C(\mathcal{B})$, so a compact collection of utilities contains arbitrarily high probability. Intersect this collection with the closed set of γ -strongly concave functions. The resulting compact collection has a common curvature bound and excludes a probability β that can be made arbitrarily small.

Compactness allows us to choose finitely many of the retained constructed utilities as representatives $V_1, \dots, V_T$, such that every retained utility differs from one of them by at most ε in uniform norm. Assign each consumer to

the first such representative, and assign the excluded consumers to one fixed representative. This defines measurable groups that are the same across contexts.

If two utilities differ by at most e and the first is γ -strongly concave, their maximizing bundles on any common convex feasible set differ by at most $2\sqrt{e/\gamma}$. Indeed, the utility loss at the second bundle is at most $2e$, whereas strong concavity bounds it below by $\gamma/2$ times the squared distance. If the omitted probability is β , the resulting uniform expected quantity error is at most

$$2\sqrt{e/\gamma} + \beta \operatorname{diam}(\mathcal{X}). \quad (\text{B.8})$$

The probabilities of the groups become the common mixture shares w_t .

B.5. Polynomial utilities and standard Gaussian shocks

The representatives are strongly concave but may not be smooth. To obtain the smooth fixed-feature utilities required by Section 3.1, we approximate them by polynomials. Each representative constructed in (B.5) can be written using a constant $\delta > 0$, constants a_r , and vectors b_r , $r = 1, \dots, n$, as $V(q) = -\delta\|q\|_2^2 + \min_{r \leq n} \{a_r + b_r^\top q\}$. For a smoothing parameter $\tau > 0$, replacing the minimum by the smooth function

$$V_\tau(q) = -\delta\|q\|_2^2 - \tau \log \sum_{r=1}^n \exp\{-(a_r + b_r^\top q)/\tau\}$$

changes its value by at most $\tau \log n$ and preserves $-\nabla^2 V_\tau \succeq 2\delta I$. On the enclosing box $\mathcal{B}$, tensor-product Bernstein polynomials approximate this smooth function uniformly together with its first and second derivatives. Choose a polynomial $\tilde{v}$ with arbitrarily small value error and Hessian error less than δ in operator norm. Then $-\nabla^2 \tilde{v} \succeq \delta I$. The utility-to-choice bound used in (B.8) makes the resulting quantity error arbitrarily small, uniformly over contexts.

Let $\tilde{v}_t$ be the finitely many resulting polynomials and $\mu_0 > 0$ their common strong-concavity margin. For $G \sim N(0, I_J)$, let $q_{t,\sigma}(c, G)$ maximize $\tilde{v}_t(q) + \sigma G^\top q$. The optimality inequalities for the perturbed and unperturbed choices imply

$$\|q_{t,\sigma}(c, G) - q_{t,0}(c)\|_2 \leq \frac{\sigma\|G\|_2}{\mu_0}. \quad (\text{B.9})$$

The perturbation therefore adds at most $\sigma\mathbb{E}\|G\|_2/\mu_0$ to the expected quantity error. Dividing the entire objective by σ leaves its maximizer unchanged. Thus $v_t = \tilde{v}_t/\sigma$ with standard Gaussian shocks gives the same choices.

Collect the monomials of these finitely many v_t 's into a common basis $\psi_1, \dots, \psi_L$. Choose $\mu > 0$ below their minimum curvature bound on $\mathcal{B}$, and define

$$\Theta = \left\{ \theta : -\sum_{\ell=1}^L \theta_\ell \nabla^2 \psi_\ell(q) \succeq \mu I \quad \text{for every } q \in \mathcal{B} \right\}. \quad (\text{B.10})$$

Each matrix restriction is affine in θ , so Θ is convex and contains every component coefficient vector. The features are smooth polynomials. This satisfies the strong-concavity condition and gives the representation in (3), with the same utilities and mixture shares across contexts.

Choose the extension perturbation, finite-record approximation, finite-type approximation, polynomial approximation and Gaussian perturbation to contribute less than $\varepsilon/5$ each. The couplings constructed at each step bound the corresponding Wasserstein distances. Their triangle inequality yields $\sup_{c \in \mathcal{C}} W_1(P^*(c), \hat{P}(c)) < \varepsilon$, which proves Theorem 2. $\square$

Appendix C: Numerical Implementation

C.1. Computational Performance

C.1.1. Instance Generation For each instance in Table 1, we independently generate utility coefficients and consumer observations as follows.

For $J = 4$ and 10 , we generate a dense positive-definite matrix

$$B = Q \operatorname{diag}(\lambda_1, \dots, \lambda_J) Q^\top,$$

where Q is obtained from the QR decomposition of a matrix of independent standard Gaussian entries. The eigenvalues are geometrically spaced between 1 and 20 and then rescaled to sum to J .

For $J = 50$, we first generate a Gaussian matrix G , form $W = (G + G^\top)/2$, and set $W_{jk} = 0$ whenever $|j - k| > 2$. Writing $\lambda_{\min}$ and $\lambda_{\max}$ for the extreme eigenvalues of W , we set

$$\tau = \frac{\lambda_{\max} - \lambda_{\min}}{19}, \quad C = W - \lambda_{\min} I_J + \tau I_J, \quad B = \frac{J}{\operatorname{tr}(C)} C.$$

The diagonal shift ensures positive definiteness while preserving the zero interactions.

The remaining true coefficients are drawn independently as

$$a_j \sim U[1.3, 2.4], \quad \alpha_j \sim U[0.3, 0.7], \quad c_j \sim U[0.04, 0.12],$$

with outside coefficient $\eta = 0.6$. The quadratic specification sets $\alpha = c = 0$. For the nonlinear specification,

$$-\nabla^2 v_\theta(q) = B + \operatorname{diag}\left(\frac{\alpha_j}{(1 + q_j)^2} + 3c_j q_j^2\right),$$

so positive-definite B and nonnegative α, c ensure strong concavity.

For each observation, prices satisfy $p_j = \exp(U_j)$, where $U_j \sim U[-0.45, 0.45]$. We draw an availability probability $r \sim U[0.65, 1]$, include each good independently with probability r , and ensure that at least two goods are available. The budget is $m = |A|V$, where $V \sim U[0.25, 1.25]$. Independent shocks follow $N(0, I_J)$. Quantities are obtained by solving the consumer problem with linear outside utility. The quadratic and nonlinear designs with the same replication index share a, B , prices, budgets, assortments, and shocks.

All methods use the parameter domain

$$\begin{aligned} B &\succeq 10^{-5} I_J, & a_j &\in [0, 6], & B_{jj} &\in [0.001, 4], \\ B_{jk} &\in [-4, 4] & & \text{for permitted interactions,} \\ \alpha_j &\in [0, 1.5], & c_j &\in [0, 0.5], & \eta &\in [0, 2]. \end{aligned}$$

Structurally absent interactions remain zero. The common initialization is

$$a_j = 1.8, \quad B = I_J, \quad \alpha_j = 0.4, \quad c_j = 0.08, \quad \eta = 0.4,$$

with α, c omitted for quadratic utility. The quadratic and nonlinear specifications contain, respectively, 15 and 23 parameters for four goods, 66 and 86 for ten goods, and 198 and 298 for fifty goods.

C.1.2. Estimation Methods The three likelihood implementations share the same likelihood, face-specific Jacobian, parameter domain, and outer optimization procedure. They differ in the evaluation of the Gaussian probability term and its derivatives.

The Dutt implementation uses characteristic-function inversion with a sum over subsets and tensor Gauss–Hermite quadrature. It starts with eight positive-axis nodes and increases this number to 16 and 32 when required by the stopping checks below. Derivatives are obtained by differentiating the same quadrature approximation.

The GHK implementation uses sequential truncated-normal simulation with scrambled Sobol points. Simulation points are fixed during optimization, and the variable ordering is determined from the observed geometry and common initialization. We differentiate the simulator analytically using reverse-mode differentiation. The simulation sizes are 256, 1,024, 4,096, and 16,384 draws.

Our implementation evaluates the likelihood and its analytic gradient together, using the same nodes in the one-dimensional integral. It starts with 64 Gauss–Legendre nodes and increases the number to 128, 256, and 512 when necessary. Observations whose likelihood does not require integration are evaluated directly. The fifty-good implementation exploits the banded interaction structure in matrix factorizations and solves.

For optimal transport, let

$$U_\beta(q, y) = v_\theta(q) + \eta y,$$

and let K_i denote the feasible consumption set under observation i 's prices, budget, and assortment. Adapting the conditional dual formulation of Carlier et al. (2016), we minimize

$$\hat{F}_{\text{OT}}(\beta) = \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{R} \sum_{r=1}^R \max_{(q, y) \in K_i} \{U_\beta(q, y) + (\epsilon^{(r)})^\top q\} - U_\beta(q_i, y_i) \right],$$

where $y_i = m_i - p_i^\top q_i$. We use $R = 64$ fixed Gaussian draws constructed from 32 scrambled Sobol points and their antithetic counterparts. The same draws are reused across observations, while each observation retains its own feasible set K_i .

Because U_β is linear in β , write $U_\beta(q, y) = \beta^\top \psi(q, y)$. The feature vector contains quantities, quadratic products, logarithmic and quartic terms where applicable, and outside consumption. The gradient is

$$\nabla \hat{F}_{\text{OT}}(\beta) = \frac{1}{nR} \sum_{i=1}^n \sum_{r=1}^R \psi(q_{ir}^*, y_{ir}^*) - \frac{1}{n} \sum_{i=1}^n \psi(q_i, y_i),$$

where (q_{ir}^*, y_{ir}^*) solves the simulated consumer problem. Thus, every objective and gradient evaluation requires simulated optimal choices. Quadratic choices are computed by an active-set quadratic-programming routine; nonlinear choices use a projected-gradient routine with backtracking. Previous choices are reused as starting points when memory permits.

We implement the SDP direction calculation in Section 5.1 using CVXPY and CLARABEL. The BFGS matrix starts at the identity and uses Powell damping. An Armijo line search starts with step size one and halves it until a feasible decrease is obtained.

C.1.3. Computing Environment and Stopping Criteria Computations use an Intel Core i5-13400F processor with 10 physical cores, 16 logical processors, and approximately 32 GB of memory, running Windows and Python 3.10. Up to four estimation jobs run concurrently. Each job uses one computational thread. For the computational comparison in Table 1, every fit receives the same 600-second wall-clock limit and a 2 GiB worker-memory limit.

Fitting time includes method-specific setup and all optimization steps, including integration or simulation refinement and the consumer optimizations required by optimal transport. Data generation, common data loading, and independent final evaluation are excluded. The fifty-good Dutt configurations are omitted under the memory limit and reported as NA.

We check stationarity over the original parameter domain Θ using

$$r(\hat{\beta}) = \left\| \hat{\beta} - \Pi_{\Theta} \left(\hat{\beta} - \nabla F(\hat{\beta}) \right) \right\|_{\infty},$$

where Π_{Θ} is Euclidean projection onto Θ . A completed estimation must satisfy the parameter constraints and have residual below 10^{-4} . For MLE, the stopping rule additionally requires the average negative log likelihood to agree with a 192-node reference calculation within 10^{-6} . Independent evaluations using 128 and 192 nodes, or 192 and 256 nodes when necessary, check that objective and gradient discrepancies are below 10^{-6} . Resolution increases use the remaining portion of the original 600-second budget. Optimal transport is checked using the gradient of its own simulated objective.

Table EC.1 reports the number of estimations that satisfy these stopping criteria within the resource limits.

Table EC.1 Completed estimations out of ten replications

Goods	Utility	Dutt MLE	GHK MLE	One-dimensional MLE	Optimal transport
4	Quadratic	10/10	10/10	10/10	10/10
4	Nonlinear	10/10	10/10	10/10	7/10
10	Quadratic	0/10	4/10	10/10	10/10
10	Nonlinear	0/10	7/10	10/10	0/10
50	Quadratic	0/10	0/10	10/10	0/10
50	Nonlinear	0/10	0/10	10/10	0/10

C.2. Empirical Performance

C.2.1. Data Construction Table EC.2 summarizes the estimation samples for CJ and KuaiRand. We randomly assign approximately 80% of households or users to training and the remainder to testing.

Table EC.2 Empirical samples

	Complete Journey	KuaiRand
Training observations	53,769	22,275
Test observations	13,212	5,555
Training households/users	1,972	800
Test households/users	494	200
Median available goods	26	31
Median consumed goods	7	29

Note: Medians use the combined sample.

Complete Journey. We aggregate household purchases by week into 29 categories using the dataset’s department and product classifications. We exclude coupon-bookkeeping and gift-card transactions and household-weeks with negative category expenditure or nonpositive total expenditure. Quantities are category expenditures in dollars, prices are one, and the budget is total weekly expenditure. All observations exhaust this budget. We approximate availability by the union of categories sold that week at the stores the household visited.

KuaiRand. We aggregate the KuaiRand-1K playback records from April 8 to May 8, 2022 by user, day, and video category (Gao et al. 2022). We use the 38 observed first-level categories in the supplementary category file and retain unclassified videos as a 39th category.² After removing duplicate records, quantities are total viewing minutes by category. The assortment contains all categories exposed that day, including those with zero viewing time. Prices are one, and the daily budget is 1,440 minutes (24 hours). We retain user-days with positive total viewing time. All retained observations leave time unspent.

C.2.2. Model Specifications Estimation uses ten-dollar units for CJ expenditure and minutes for KuaiRand viewing time. The KT models use independent standard Gaussian shocks and deterministic utility

$$v_{\theta}(q) = a^{\top} q - \frac{1}{2} q^{\top} B q + \sum_{j=1}^J \sum_{\ell=1}^3 \alpha_{j\ell} \left[s_{j\ell} \log \left(1 + \frac{q_j}{s_{j\ell}} \right) - q_j \right].$$

Quadratic sets $\alpha = 0$; Quadratic + log estimates $\alpha_{j\ell} \geq 0$, with $(s_{j1}, s_{j2}, s_{j3}) = (s_j/4, s_j, 4s_j)$ and s_j fixed at the median positive training quantity. Both models allow all pairwise interactions. We impose $B \succeq 10^{-5} I_J$, $|a_j| \leq 100$, $|B_{jk}| \leq 100$, and $0 \leq \alpha_{j\ell} \leq 100$.

The MDCEV benchmark uses the translated-power utility of Bhat (2008). We estimate good-specific intercepts and translation parameters and a common curvature parameter, with independent unit-scale extreme-value shocks. For KuaiRand, MDCEV includes an always-available outside good with quantity $q_0 = 1440 - \sum_j q_j$, unit price, its own translation parameter, and an intercept normalized to zero. The KT models use the inequality budget without outside utility. Mixture likelihoods are formed at the household-week or user-day level.

C.2.3. Estimation and Computing Times KT models use the constrained BFGS procedure in Section 5.1; MDCEV uses its closed-form likelihood and L-BFGS-B. Four-type models are estimated by generalized EM, initialized from the corresponding single-type estimates. Empirical fits run concurrently, with one computational thread per fit, on the hardware described in Section C.1.3. Each fit uses a one-hour wall-clock limit and stops when the absolute change in average training NLL over five iterations is at most 10^{-5} . For mixtures, this criterion uses the observed mixture likelihood after each EM cycle.

For binding-budget observations, the KT likelihood integrates over $\lambda \geq 0$; numerical integrals use 80-node Gauss–Legendre quadrature with adaptively chosen bounds. Slack-budget observations use $\lambda = 0$ and require no numerical integration.

Reported likelihoods use dollar units for CJ and minute units for KuaiRand. Within each dataset, all models use the same density measure: Euclidean volume on the observed budget face for CJ, and the positive viewing-quantity coordinates for KuaiRand.

² <https://zenodo.org/records/18159199>

Table EC.3 Fitting time in minutes for the empirical models

Model	Complete Journey	KuaiRand
MDCEV	0.42	0.32
Quadratic	24.25	2.80
Quadratic + log	60.00 [†]	10.89
MDCEV, four types	17.57	9.23
Quadratic, four types	60.00 [†]	31.43
Quadratic + log, four types	60.00 [†]	60.00 [†]

Notes: Times report each fit and exclude data preparation, final scoring, and the single-type fits used to initialize mixtures. A dagger marks a fit that reached the one-hour limit before meeting the NLL stopping criterion.